N° d'ordre : 2014-19-TH

SUPELEC

ECOLE DOCTORALE STITS
« Sciences et Technologies de l'Information des Télécommunications et des Systèmes »

THÈSE DE DOCTORAT

DOMAINE : STIC
Spécialité : Télécommunications

Soutenue le
7 octobre 2014
par :
Andrés O. ALTIERI

On Large Cooperative Wireless Network Modeling through a Stochastic Geometry Approach

Modélisation de Réseaux sans Fils de Grandes Dimensions à l'aide de la Géométrie Stochastique

Directeurs de thèse :	Cecilia G. GALARZA	Associate Professor, Univ. of Buenos Aires, Arg.
	Hikmet SARI	Professor, Supelec, France
Encadrants de thèse :	Pablo PIANTANIDA	Assistant Professor, Supelec, France
	Leonardo REY VEGA	Assistant Professor, Univ. of Buenos Aires, Arg.
Composition du jury :		
Rapporteurs :	Bartłomiej BŁASZCZYSZYN	Research Director, INRIA, France
	Martin HAENGGI	Professor, Notre Dame University, United States
Examinateurs :	Laurent DECREUSEFOND	Professor, Telecom ParisTech, France
	Pierre DUHAMEL	Research Director, CNRS, France
	Aris MOUSTAKAS	Assistant Professor, Athens University, Greece
	Pablo PIANTANIDA	Assistant Professor, Supelec, France
	Hikmet SARI	Professor, Supelec, France

To my parents

Abstract

The main goal of this work is to study cooperative aspects of large wireless networks from the perspective of stochastic geometry. This allows the consideration of important effects such as the random spatial distribution of nodes, as well as the effects of interference and interference correlation at receivers, which are not possible when a single link is considered in isolation.

First, some aspects of the performance of the relay channel in the context of a large wireless network are considered. Mainly, the performance, in terms of outage probability (OP), of a single full-duplex relay channel utilizing decode-and-forward (DF) or compress-and-forward, when the interference is generated by uniform spatial deployment of nodes, modeled as a Poisson point process. The OP performance of these two protocols is compared with a point-to-point transmission and with a half-duplex DF protocol. Afterwards, the case in which more than one transmitter in the network may use a relay is considered. The effects of cooperation versus interference are studied, when the users use either full-duplex DF, or point-to-point transmissions. DF is chosen, i.e. relays are activated, in a random independent manner. In typical operating conditions, the best policy is found to be to either activate all the relays at the same time, or none at all, depending on networks setup parameters, such as the source-destination distance and average source-relay distance. This is condition is shown not to be valid when the OP is large and outside of the typical operating range.

In a second phase, this work explores the advantages that could be obtained through out-of-band device-to-device (D2D) video file exchanges in cellular networks. These advantages are measured in terms of the fraction of requests that can be served in a time-block through D2D, thus avoiding a downlink file transfer from the base station. For this, a stochastic geometry framework is introduced, in which the user file-caching policy, user pairing strategy, and link quality and scheduling issues are considered. This model shows that, even through simple strategies, D2D exchanges show promising gains.

Publications

Accepted journal papers

- A. Altieri, L. Rey Vega, P. Piantanida, C. G. Galarza, "On the Outage Probability of the Full-Duplex Interference-Limited Relay Channel". *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 9, pp. 1765,1777, Sept. 2014. doi: 10.1109/JSAC.2014.2330192
  `http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6832447`.

- A. Altieri, L. Rey Vega, P. Piantanida, C. G. Galarza, "Analysis of a Cooperative Strategy for a Large Decentralized Wireless Network". *IEEE/ACM Transactions on Networking*, vol. 22, no. 4, pp. 1039–1051, Aug. 2014 doi: 10.1109/TNET.2013.2269054
  `http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6552219`.

Submitted journal papers

- A. Altieri, P. Piantanida, L. Rey Vega, C. Galarza, "On Fundamental Trade-offs of Device-to-Device Communications in Large Wireless Networks". Submitted to the *IEEE Transactions on Wireless Communications*, May 2014.
  Available at `http://arxiv.org/abs/1405.2295`.

Accepted conference papers

- A. Altieri, L. Rey Vega, P. Piantanida, C. G. Galarza, "A Stochastic Geometry Approach to Distributed Caching in Large Wireless Networks". *IEEE Symposium on Wireless Communication Systems (ISWCS)*, 26-29 August 2014, Barcelona, Spain.

- A. Altieri, L. Rey Vega, P. Piantanida, C. G. Galarza, "Cooperative Unicasting in Large Wireless Networks". *IEEE Symposium on Information Theory (ISIT)*, 7-12 July 2013,

Istanbul, Turkey.

- A. Altieri, L. Rey Vega, P. Piantanida, C. G. Galarza, "On the Balance Between Cooperation and Interference in Dense Wireless Networks". *IEEE Symposium on Wireless Communication Systems (ISWCS)*, 28-31 Aug. 2012 , Paris, France.

- A. Altieri, L. Rey Vega, P. Piantanida, C. G. Galarza, "Cooperation versus Interference in Large Wireless Relay Networks". *IEEE Symposium on Information Theory*, 1-6 July 2012, Cambridge (MA), United States.

- A. Altieri, L. Rey Vega, P. Piantanida, C. G. Galarza, "Cooperative Strategies for Interference-Limited Wireless Networks". *IEEE Symposium on Information Theory*, July 31 2011-Aug. 5 2011, St. Petersburg, Russia.

Acknowledgments

I would like thank my advisors and co-advisors for their support and guidance throughout my PhD. Each of them played a different and important role through the various stages of my work. I would like to thank my advisor in Argentina, Dr. Cecilia Galarza who gave me complete freedom, thoughtful advice, and assistance whenever I needed it. Also, I would like to thank my advisor in France Dr. Hikmet Sari for accepting me and ensuring smooth stays at the Telecommunications Department at Supelec. I would like to thank my co-advisor Dr. Pablo Piantanida for his support and guidance, specially during my stays at Supelec, for always encouraging me to pursue new goals and work hard. Finally, I would like to thank my co-advisor Dr. Leonardo Rey Vega for his guidance and support, specially during the initial stages of my work.

I would like to thank my friends who listened, encouraged and helped me throughout my work. I want to thank Ovi, for all the fun moments, conversations, and support both at the lab and outside. To Germán, for all the good times in France, at Supelec and when travelling, and for all his help while I was in Argentina. To all the friends I made during my stays at Supelec: Chao, Chienchun, Myriam, Kareem, Zheng, Matha, and the others.

I want to thank my parents for all the effort and care they put into my upbringing. They have unconditionally supported me, and encouraged me to do my best, teaching me by example. Finally, I want to thank Laura, for always being there for me, for encouraging me and caring for me when I needed it. Also, for supporting me through those late work nights and times we were apart.

Contents

List of Figures

Abbreviations

AWGN Additive white Gaussian noise

BS Base Station

CCSG Complex circularly symmetric Gaussian

CF Compress-and-Forward

CSI Channel state information

D2D Device-to-device communications

DF Decode-and-Forward

DT Direct/point-to-point transmission

OP Outage probability

PPP Poisson point process

SINR Signal-to-interference-noise ratio

SIR Signal-to-interference ratio

SNR Signal-to-noise ratio

TDMA Time-division multiple access

Mathematical symbols

$(\cdot)^*$ Complex conjugation

$(\cdot)^c$ Complement of a set

$F_X(\cdot)$ Cumulative distribution function of the random variable X

$I(X,Y)$ Mutual information between the random variables X and Y

$O(\cdot)$ Big O notation: $f(x) = O\left(g(x)\right)$ as $x \to x_0$ if there exists $M > 0$ such that $|f(x)| \leq M|g(x)|$ in some neighborhood of x_0

$\mathcal{B}(x,y)$ In $\mathbb{R}^d$, ball of radius y centered at x

$\mathcal{B}_{\mathcal{X}}$ Borel σ-algebra induced by a suitable topology on a set $\mathcal{X}$

$\mathsf{C}(\cdot)$ The Shannon capacity function $\mathsf{C}(x) = \log_2(1+x)$, for $x \geq 0$

$\Re(\cdot)$ Real part of a complex number

$\bar{F}_X(\cdot)$ Complementary cumulative distribution function of the random variable X

$\mathbb{1}\{A\}, \mathbb{1}_A$ Indicator of an event A

$\setminus$ Set substraction $A \setminus B = A \cap B^c$

$\{A_i\}_{i\in\mathcal{I}}$ A family of objects, indexed by a set $\mathcal{I}$; when clear from context, $\mathcal{I}$ is omitted

$||\cdot||$ Euclidean norm in $\mathbb{R}^d$

Chapitre 1

Modélisation de Réseaux sans Fil de Grandes Dimensions à l'aide de la Géométrie Stochastique

1.1 Introduction

Les communications sans fil font partie d'un segment qui se développe très rapidement dans l'industrie des télécommunications. Nouvelles applications sont envisagées régulièrement par les spécialistes, et la tendance est de concevoir un schéma de connexion où tous les appareils électroniques sur la terre seront reliés à un réseau. Ces applications couvrent les domaines les plus divers comme les communications entre machines (*machine-to-machine)*, les réseaux électriques intelligents (*smart-grids)*, réseaux de capteurs sans fil (*wireless sensor networks*), les réseaux de sécurité publique, etc..

Les communications cellulaires ont expérimenté une des croissances les plus fortes, notamment en raison d'un changement dans le type de trafic. Celui-là, qui était originalement formé par transmission de voix, a changé dernièrement à un trafic du type IP, grâce à l'apparition des téléphones intelligents (*smartphones*), tablettes et *streaming* de vidéo [1]. Les nouvelles exigences représentent des défis importants pour l'ingénierie des communications, ayant besoin de nouvelles et importantes progrès dans le domaine. Dans ce scénario, il est nécessaire de définir de nouveaux paradigmes, architectures, et interactions entre les nœuds des réseaux, pour maximiser l'utilisation des ressources et l'efficacité spectrale. La coopération entre nœuds, va donc jouer un rôle important dans cette tâche et les aspects principaux devraient être étudiés à l'aide de modèles mathématiques qui synthétisent les éléments essentiels des réseaux sans fil.

La géométrie stochastique [2] et, particulièrement, la théorie des processus ponctuels sont des outils mathématiques spécialement adaptés pour modéliser les aspects principaux des réseaux sans fils, où les nœuds sont aléatoirement distribués dans l'espace. L'objectif général de ce travail est l'analyse de certains aspects des réseaux sans fils coopératifs en utilisant des processus ponctuels. Ce chapitre contient un résumé des résultats principales de la thèse, qui son développés en détail dans les chapitres suivants en anglais.

1.2 Modèles de canaux sans fil considérés

Le milieu de propagation électromagnétique introduit des difficultés importantes pour l'établissement des communications sans fils. Les transmissions sans fils sont, presque toujours, omnidirectionnelles, et ça signifie que, en plus de l'atténuation à cause de la distance, il existent d'autres phénomènes d'atténuation et distorsion, associés à la géométrie de l'espace que les ondes électromagnétiques traversent. Nous pouvons nommer trois phénomènes principaux [3] :

- L'atténuation à cause de la distance parcouru par l'onde électromagnétique (*path loss*).
- L'atténuation à cause des objets de grandes dimensions qui se trouve dans l'espace (*shadow-fading*).
- L'atténuation à cause de la propagation des ondes par une grande quantité de chemins avec des délais de propagation différents (*multipath fading* ou *fading* simplement).

Normalement, ce n'est pas possible de prévoir l'atténuation à cause de la géométrie de l'espace, et donc des modèles aléatoires sont considérés. Pour l'atténuation à cause de la distance, des modèles déterministes et paramétriques sont souvent employés.

Généralement, la modélisation des canaux sans fils suppose que les canaux sont linéaires et variants dans le temps, caractérisés par un réponse $c(t, \xi)$. Celle-ci représente la réponse du canal dans l'instant t à une impulsion qui fut à l'entrée du canal au temps $t-\xi$. Normalement, cette réponse est modelée à l'aide d'un processus stochastique. L'étude de la fonction d'autocorrelation du processus, en considérant des hypothèses raisonnables, permet d'introduire deux paramètres qui caractérisent le comportement du canal. Ces paramètres sont :

- La dispersion par chemins multiples (*multipath spread*) T_m, qui donne une estimation du temps dans lequel la sortie du canal est décorrélée avec l'entrée. L'inverse de ce paramètre, est la largeur de bande de cohérence, (*coherence bandwidth*) B_{coh}. On dit qu'un canal passe-bande avec bande passante W est non-sélectif en fréquence (*flat fading* ou *frequency non-selective*) si $W \ll B_{\text{coh}}$. Dans ce cas, toutes les fréquences à l'entrée du canal seront affectés de la même manière.

- La dispersion Doppler du canal (*Doppler spread*) B_d qui indique la vitesse de changement du canal dans le temps. L'inverse de ce paramètre est appelé le temps de cohérence du canal (*coherence time*), T_{coh}. On dit que le canal est à évanouissement lent (*slow fading*) si $T_s \ll T_{\text{coh}}$ où T_s indique la durée d'un bloc.

Dans cette thèse, nous concentrons notre attention sur les canaux non-sélectifs en fréquence et à évanouissement lent ; comme conséquence le canal équivalent en bande de base peut être écrit de la manière suivante :

$$y(t) = gx(t), \tag{1.1}$$

où $x(t)$ est le signal transmis, $y(t)$ le signal reçu, et g est la réponse scalaire complexe du canal. La valeur de g est choisie en fonction d'un modèle qui considère les phénomènes d'atténuation du canal sans fils que nous avons signalé ci-dessus. En général, tous les phénomènes ne sont pas considérés simultanément. Un modèle standard considère seulement la présence de *path loss* et de *fading*, en ignorant le *shadow fading*. Dans ce cas, si l'émetteur est placé en u et le récepteur est placé en v, le gain du canal peut être écrit comme suit :

$$g = l(u, v)^{\frac{1}{2}} h, \tag{1.2}$$

où $l(u, v)$ est une fonction qui représente le *path loss* entre u et v, et h est une constante complexe qui contient le *fading*. Normalement la fonction $l(u, v)$ dépend seulement de la distance entre les points, c'est à dire, $l(u, v) \equiv l(||u - v||)$. Il existe de nombreux exemples de ces fonctions, la plus habituelle est $l(u, v) = ||u-v||^{-\frac{\alpha}{2}}$, où α est l'exposant d' atténuation avec la distance, normalement dans l'intervalle $(2,\ 6)$. Dans ce travail nous considérons $l(u, v) = ||u - v||^{-\frac{\alpha}{2}}$. L'hypothèse sur évanouissement lent implique que le coefficient sera fixé avant de commencer la transmission et il restera constant pendant la transmission. Il existe de nombreux modelés statistiques pour sélectionner le coefficient h. Le modèle le plus habituel est le modèle *fading* Rayleigh. Dans ce modèle, $|h|$ es une variable aléatoire Rayleigh et la phase de h est uniforme en $[0, 2\pi)$. Comme conséquence $|h|^2$ a une distribution exponentielle. Pendant ce travail, nous employons en général le modèle de *fading* Rayleigh.

Dans notre contexte, les modèles sont toujours en temps discret, pouvant généralement être transformés en modèles en temps continu sans perte par une procédure d'échantillonnage (voir [4]). Ainsi donc, le modèle à évanouissement lent que nous employons pour cette thèse est :

$$y_i = l(u, v)^{-\frac{\alpha}{2}} h x_i, \tag{1.3}$$

où $i = 1, 2, \ldots$ indique le temps. Le développement de ces modèles est complété plus en détail dans l'Annexe A.

1.3 Mesures de capacité pour canaux *slow fading*

L'un des principaux objectifs de l'étude des systèmes de communication est de connaître les taux maximaux qui peuvent être transmis de manière fiable sur un canal de communication. La première formulation rigoureuse du problème et les premiers résultats ont été développés par C. E. Shannon en 1948 [5], et ils ont conduit à la formulation de la Théorie de l'Information qui compte aujourd'hui de nombreuses applications dans des domaines divers [4].

Les premiers résultats de Shannon ont été orientés à l'étude des canaux point à point, c'est-à-dire, entre un émetteur et un récepteur seulement (Fig. 1.1). Les principaux composants du canal sont les suivants :

- Un alphabet de symboles $\mathcal{X}$ qui peuvent être transmis à travers le canal.
- Un alphabet de symboles $\mathcal{Y}$ qui peuvent être observés à la sortie du canal.
- Une distribution de probabilité conditionnelle $p_{Y^n|X^n}$ qui s'utilise pour générer une séquence $Y^n \in \mathcal{Y}^n$ de longueur n à la sortie du canal à partir d' une séquence $X^n \in \mathcal{X}^n$ de longueur n a la entré du canal.

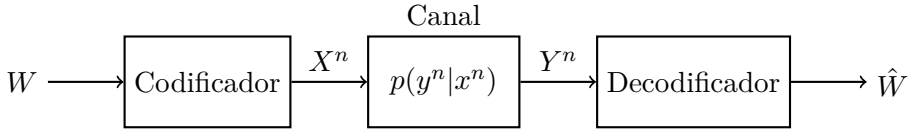

FIGURE 1.1: Schéma fonctionnel d'un canal point à point.

Pour effectuer une transmission de longueur n dans le canal, on utilisera un code (M, n) constitué comme suit :

- Un ensemble de messages $\{1, \ldots, M\}$ dont l'un, W, est choisi pour être transmis.
- À l' émetteur : une fonction de codage $X^n : \{1, \ldots, M\} \to \mathcal{X}^n$, qui assigne à chaque message une séquence de longueur n qui sera transmise dans le canal.
- Au récepteur : une fonction de décodage $Y^n : \mathcal{Y}^n \to \{1, \ldots, M\}$ qui génère une estimation $\hat{W}$ du message transmis.

On définit le taux R d'un code (M, n) (en bits/symbole transmis) par :

$$R = \frac{\log_2(M)}{n}. \tag{1.4}$$

La probabilité d'erreur associé à un mot de code est défini comme :

$$P_{e,i} = \mathbb{P}(\hat{W} \neq W | W = i). \tag{1.5}$$

On dit que le taux R est atteignable s'il existe une suite de codes $(2^{nR}, n)$ qui vérifient :

$$\max_{i=1,\ldots,2^{nR}} P_{e,i} \underset{n\to\infty}{\longrightarrow} 0. \tag{1.6}$$

La capacité du canal est définie par :

$$C = \sup\{R : R \text{ est atteignable}\}. \tag{1.7}$$

Shannon a été un pionnier qui a démontré qu'on peut atteindre une probabilité d'erreur aussi petite que l'on veut avec un taux strictement positif. En plus, sa recherche montra que les taux utilisés en son temps étaient bien inférieurs à ceux qui pourraient être utilisés [3].

Le modèle le plus classique pour représenter les canaux de communications pratiques est le canal à bruit blanc additif gaussien (Fig. 1.2) (*additive white Gaussian noise*, AWGN) dont la relation entrée-sortie est :

$$Y_i = gX_i + Z_i, \quad i = 1, 2, \ldots \tag{1.8}$$

où $\{Z_i\}_i$ est une suite de variables aléatoires indépendantes, identiquement distribuées normales complexes, circulaires, symétriques [6] avec variance σ^2, indépendantes de $\{X_i\}_i$, et g coefficient fixé, complexe et connu au moins à la destination. Ce canal est utilisé souvent

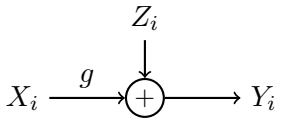

FIGURE 1.2: Représentation du canal à bruit blanc additif gaussien obtenu par la équation (1.8).

pour représenter un canal idéal avec bruit thermique dans le récepteur de la destination. Les alphabets d'entrée et sortie de ce canal sont continus. Si nous utilisons un code (M, n) arbitraire comme ci-dessus, nous pouvons choisir une suite de symboles suffisamment éloignés de telle sorte que la probabilité d'erreur soit aussi petite que l'on veut, en étant la capacité infinie. Donc, une restriction additionnel dans les codes doit être incorporé. En général, si $(x_1, \ldots, x_n)$ désigne une suite d'entrée, on utilise une restriction d'énergie comme suit :

$$\frac{1}{n}\sum_{i=1}^{n} |x_i|^2 \leq P. \tag{1.9}$$

Shannon a démontré que la capacité de ce canal est :

$$C = \sup_{X:\mathbb{E}[|X|^2]<P} I(X,Y) = \log(1+\text{SNR}), \tag{1.10}$$

où $\text{SNR} = |g|^2P/\sigma^2$ est le rapport signal sur bruit à la destination et $I(X,Y)$ est l'information mutuelle entre les variables X et Y définie comme suit :

$$I(X,Y) = \mathbb{E}\left[\log\left(\frac{f_{X,Y}(x,y)}{f_X(x)f_Y(y)}\right)\right]. \tag{1.11}$$

où f_X, f_Y, $f_{X,Y}$ sont les densités des variables indiquées. Le supremum en (1.10) est dans l'ensemble de tous les densités de probabilité qui satisfassent la restriction sur la variance, et il est atteignable si $X \sim \mathcal{CN}(0,P)$. En plus, la démonstration du théorème montre que la capacité est atteignable si les mots des codes sont générés comme une suite de variables aléatoires gaussiennes indépendantes avec variance P.

Dans le canal (1.3) introduit précédemment, on observe que pour chaque réalisation du coefficient de *fading* g, le canal (1.3) correspond au canal gaussien (1.8). En conséquence, la capacité est liguée à la réalisation du *fading*.

$$C(g) = \log\left(1+\frac{|g|^2P}{\sigma^2}\right). \tag{1.12}$$

Comme le coefficient de *fading* est aléatoire et il est fixé avant de commencer la transmission, la capacité du canal sera aussi un variable aléatoire : avant de transmettre, le *fading* est sélectionné et il définit la capacité du canal pendant la transmission. Si la notion traditionnel de capacité est utilisée sur le canal, on devrait trouver un taux qui pourrait être transmis sans erreur sur tous les réalisations du *fading*. Comme, en général, le *fading* peut être arbitrairement petit (selon sa distribution), il y a une probabilité positif d' obtenir $C(g) < R$ pour tous les taux $R > 0$ sélectionnés. En conséquence, selon cette notion, la capacité du canal *slow fading* est égale à zéro. Ceci motive une définition différente pour la capacité de ces canaux. Si on suppose que la réponse du canal n'est pas connue à l'émetteur, il ne peut pas adapter le taux de transmission à la réalisation du canal. Ainsi, pour chaque taux R sélectionné, s' il y arrive que $\{C(g) < R\}$, le taux ne sera pas atteignable et nous dirons que il y a un événement de défaut ou *outage*. La probabilité de *outage* peut être obtenue comme suit :

$$P_{\text{out}}(g) \triangleq \mathbb{P}(C(g) < R) = \mathbb{P}(|g|^2 < \sigma^2(2^R-1)/P) = F_{|g|^2}\left(\frac{\sigma^2}{P}(2^R-1)\right), \tag{1.13}$$

qui est la probabilité d' avoir que le taux sélectionné R ne puisse pas être transmis à travers le canal. Cette probabilité motive la définition de capacité de outage du canal comme suit :

$$C_{\text{out}} = \max_{g:P_{\text{out}(g)<\epsilon}} C(g). \tag{1.14}$$

Dans le sections suivantes, nous observerons que la notion de probabilité de outage peut être généralisée à canaux slow *fading* plus générales. En plus, la probabilité de outage est un borne supérieure de la probabilité d'erreur du canal et, par conséquence, elle est une métrique importante.

1.4 Application de la géométrie stochastique à la modélisation de réseaux sans fil

Les travaux postérieurs à Shannon ont utilisé ses idées pour analyser et comprendre des scénarios plus complexes, comme le canal à accès multiple (*multiple-access channel*), le canal *broadcast*, ou le canal d'interférence [4]. Bien que l'analyse de réseaux avec nombreux nœuds est possible du point de vue de la théorie de l'information, les expressions obtenues sont, souvent, très complexes ou difficiles à interpréter et analyser. Comme les réseaux sans fils seront composés en général de nombreux utilisateurs, il est nécessaire de développer des modèles appropriés, qui permettent d'obtenir des conclusions pertinentes. Tel que discuté précédemment, le support sans fil est particulièrement défavorable à la communication, par rapport aux médias branchés : les signaux transmis subissent une atténuation avec la distance, ainsi qu'une atténuation à cause de la géométrie, souvent imprévisible, de l'environnement où les transmissions sont faites. En outre, en général il y a de nombreuses transmissions simultanées, et comme elles sont omnidirectionnelles et le spectre radioélectrique est limité, l'interférence entre les transmissions est inévitable. Dans de cas nombreux, cette interférence va être plus importante que la présence du bruit.

Considerons donc un grand réseau avec l émetteurs et m récepteurs qui partagent le même espace et le même spectre, et dont les canaux sont d'évanouissement lent comme (1.3). Ce réseau, à chaque instant peut être représenté par un système d'équations comme suit :

$$y_k = \sum_{j=1}^{l} a_{j,k} x_j + z_k, \qquad k = 1, \ldots, m, \tag{1.15}$$

où y_k est le symbole reçu par l'utilisateur k, $a_{j,k}$ est la réponse complexe du canal, x_j est le symbole émis par l'émetteur j, et z_k est le bruit complexe gaussien au récepteur. Pour étudier des réseaux sans fils, il est normalement nécessaire de considérer un grand nombre de nœuds (m et n grands), dont les positions sont généralement imprévisibles. L'analyse d'une configuration particulière du réseau ne permettra pas obtenir des conclusions sur le comportement moyen des utilisateurs surtout parce que dans une configuration donnée les canaux seront tous différents. De plus, le comportement de l'interférence est différent du comportement du bruit, car le bruit n'est pas affectée par les actions des utilisateurs, tandis

que dans le cas d'interférence, les actions de les utilisateurs qui devraient leur être bénéfique individuellement, peuvent être contre-productif si elles sont effectuées par tous les utilisateurs du réseau en même temps. Cet effet n'est pas présent quand il y seulement du bruit décorrélé aux récepteurs.

La théorie des processus ponctuels est une branche de la géométrie stochastique [2] dédiée à l'étude des distributions aléatoires de points dans l'espace. Cette théorie fournit un cadre naturel pour la modélisation et l'analyse de grands réseaux sans fil où l'interférence est une des restrictions principales dans l' étude de la performance [7, 8, 9]. En utilisant cette théorie, la plupart des phénomènes caractéristiques d'un réseau sans fil peuvent être modélisés, et les effets macroscopiques de l'interférence et sa corrélation espatio-temporelle peuvent être étudiés.

Avec cette théorie, les aspects principaux d'un réseau sans fil peuvent être modelés et, en moyennant sur toutes les configurations spatiales des points, ont peut étudier les effets macroscopiques de l'interférence et sa corrélation spatio-temporelle. En plus, la modélisation du réseau en utilisant des processus stationnaires implique que, en moyenne, la performance de tous les nœuds sera la même. Ainsi, on peut caractériser la performance globale du réseau en étudiant un nœud et obtenir des expressions fermées et compactes qui représentent tout le réseau.

Un processus ponctuel est structurellement analogue à une variable aléatoire, sauf que le résultat du processus est un ensemble nombrable de points dans un espace, au lieu d'un nombre réel. Dans le contexte des réseaux sans fil, les points seront généralement dans le plan, c'est-à-dire, à $\mathbb{R}^2$. En général, le processus ponctuel est dénoté comme $\Phi = \{u_1, u_2, \ldots\} = \{u_i\}$, avec $u_i \in \mathbb{R}^2$. Le processus ponctuel plus utilisé et étudié est le processus ponctuel de Poisson homogène ou stationnaire ; ce processus est utile pour représenter des situations où les nœuds sont distribués uniformément dans l'espace.

Definition 1.4.1 (Processus homogène de Poisson)**.** *On dit que Φ est un processus ponctuel de Poisson d'intensité $\lambda > 0$ en $\mathbb{R}^2$ si les deux conditions suivantes sont vérifiées :*

- *Étant donné un ensemble borné A, la quantité des points du processus en A, désignée comme $\Phi(A)$ est un variable aléatoire Poisson d'espérance $\lambda|A|$, où $|\cdot|$ désigne l'aire d' un ensemble.*
- *Si $A_1, \ldots, A_n$ sont ensembles disjoints et bornés, $\Phi(A_1), \ldots, \Phi(A_n)$ sont variables aléatoires indépendantes.*

Par conséquence de cette définition, on peut démontrer que, étant donné un région borné A, si $\Phi(A)$ est connu, les points du processus sont distribués uniformément et indépendamment

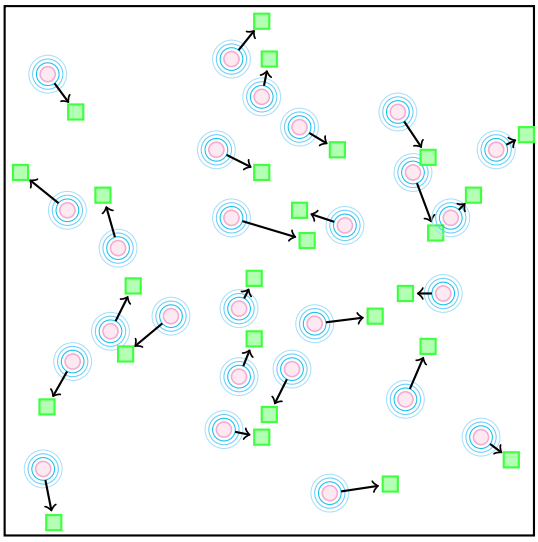

FIGURE 1.3: Réseau spatial où les émetteurs (cercles en magenta) transmettent à ses destinations (carrés en vert) par un transmission point à point.

en A. Ceci permet la simulation du processus de Poisson dans une région arbitraire ; d'abord on génère une variable Poisson d'espérance $\lambda|A|$, puis on tire le même nombre de points uniformes sur la région. L'intensité ou densité λ a unités de points par unité de surface, et elle représente la densité spatiale moyenne des nœuds du réseau.

Dans ce qui suit, nous illustrerons comment travailler avec un processus ponctuel en utilisant un exemple qui contient les difficultés et les principaux aspects du problème [10]. Supposons que dans un réseau de communications (1.15) les émetteurs sont distribués dans le plan selon un processus ponctuel stationnaire $\Phi = \{u_i\}$, et que chacun d'eux a une destination différente qui n'appartient pas au processus et avec laquelle il veut être communiquée à travers une communication point à point (Fig. 1.3). Supposons que le canal est *slow fading* comme (1.3) avec le *fading* d'espérance unitaire, et que la destination connaît le canal qui la sépare de la source, mais pas la source ne le connaît pas. Pour le moment nous ne considérons pas le bruit. Supposons que l'émetteur est situé à le origine o et que sa destination est située en d. Le signal reçu au récepteur au temps k est :

$$y_{d,k} = ||d||^{-\frac{\alpha}{2}} h_d x_{o,k} + \sum_{u \in \Phi \setminus \{o\}} ||u - d||^{-\frac{\alpha}{2}} h_{u,d} x_{u,k}, \tag{1.16}$$

où $||\cdot||$ désigne la norme euclidienne. Le premier terme est la transmission de l'utilisateur dans l'origine, et le second terme sont les transmissions d'autres utilisateurs du processus excluant l'origine. Ce canal est un canal d'interférence où plusieurs utilisateurs essayent de transmettre simultanément en générant interférences entre eux. En général, pour le décodage

au récepteur, on suppose que l'interférence est un bruit. De cette façon là, le canal est analysé comme un canal point à point au lieu de le considérer comme un canal d'interférence. Observons que l'interférence dépend de la stratégie de codage que les utilisateurs utilisent, c'est-à-dire, que l'analyse est différente et plus compliqué que le canal AWGN mentionné dans la section précédente où le bruit est statique. Cependant, nous pouvons trouver un taux atteignable si nous utilisons le même codage que dans le canal AWGN : si les mots du code des utilisateurs sont générés comme réalisations des variables gaussiennes indépendantes circulairement symétriques de variance P, l'interférence aura aussi distribution gaussienne et sa variance est :

$$I(d) = \sum_{u \in \Phi \setminus \{o\}} P||u-d||^{-\alpha}|h_{u,d}|^2. \tag{1.17}$$

Cette variance est la puissance de l'interférence en d, où les nœuds et le *fading* sont fixés. Si le *fading* et le processus ponctuel sont aléatoires, la puissance de l'interférence sera une variable aléatoire. De cette façon, la relation signal-interférence (*signal-to-interference ratio*, SIR) peut être définie comme le quotient entre la puissance d'émetteur et la puissance de l'interférence :

$$\mathrm{SIR}(o,d) = \frac{||d||^{-\alpha}|h_d|^2}{\sum_{u \in \Phi \setminus \{o\}} ||u-d||^{-\alpha}|h_{u,d}|^2}. \tag{1.18}$$

L'émetteur pourra atteindre tout taux R qui vérifie :

$$R < \log_2\left(1 + \mathrm{SIR}(o,d)\right). \tag{1.19}$$

Le problème est similaire à celui analysé pour le canal AWGN dans le sens que le taux de la droite (1.19) est une variable aléatoire, mais diffère dans le sens que, pour chaque réalisation du processus et du *fading*, ce taux n'est pas nécessairement la capacité du canal. Donc, nous pouvons appliquer ici le même critère de *outage* utilisé avant pour le canal AWGN. Comme l'émetteur n'a pas accès à la SIR au recepteur, il ne peut pas choisir R pour s' adapter au canal. On dira qu' un événement d'*outage* se produit si (1.19) n'est pas satisfait. Donc, un *outage* se produit si la réalisation du *fading* et du processus ponctuel empêchent que ces taux soient atteignables.

Pour trouver la probabilité d'*outage*, il faut calculer l'espérance de la distribution du processus en considérant seulement les réalisations du processus qui ont un point à l'origine des coordonnées. Ça c'est équivalent à faire l'espérance selon la distribution conditionnelle d'un processus avec un point à l'origine. Comme la probabilité d' avoir un processus avec un point à une coordonée spécifique dans l'espace est zéro, cette probabilité conditionnelle n'est pas définie dans le sens traditionnelle. Nous devons utiliser la théorie des distributions

de Palm pour trouver la probabilité d'*outage*, P_{out} :

$$P_{\text{out}} = \mathbb{P}^0\left(R < \log_2\left(1 + \text{SIR}(o,d)\right)\right), \tag{1.20}$$

$$= 1 - \mathbb{E}^0\left[F_{|h_d|^2}\left(\frac{(2^R-1)I(d)}{||d||^{-\alpha}}\right)\right], \tag{1.21}$$

où $F_{|h_d|^2}$ est la distribution du *fading* et $\mathbb{E}^0$ est la moyenne selon la distribution conditionnelle d'un processus qui a un point à l'origine des coordonnées mais sans considérer le point dans la moyenne. Si le *fading* est Rayleigh de moyenne unitaire, $|h_d|^2$ a une distribution exponentielle, et alors (1.21) est :

$$P_{\text{out}} = 1 - \mathbb{E}^0\left[\exp\left\{\frac{(2^R-1)I(d)}{||d||^{-\alpha}}\right\}\right] \triangleq \mathcal{L}^0_{I(d)}\left(\frac{(2^R-1)I(d)}{||d||^{-\alpha}}\right), \tag{1.22}$$

où $\mathcal{L}^0_{I(d)}$ est la transformation de Laplace de la variable aléatoire interférence, par raport à la distribution de Palm réduite du processus ponctuel. En général, les distributions de Palm du processus ponctuel sont difficiles de obtenir. Dans le processus ponctuel de Poisson sa distribution de Palm réduite coïncide avec la distribution du processus sans conditionner (Théorème 4.6)., ainsi, pour le processus de Poisson :

$$\mathcal{L}^!_{I(d),0}(\omega) = \mathcal{L}_{I(d)}(\omega). \tag{1.23}$$

En plus, dans le cas du processus de Poisson, la transformation de Laplace a une forme fermée (Lemme 4.3), donc :

$$P_{\text{out}} = 1 - \exp\left\{-\lambda\tilde{C}(2^R-1)^{2/\alpha}||d||^2\right\}, \tag{1.24}$$

où $\tilde{C}$ est donnée par (5.10). En fait, la stationnarité du processus implique que cette probabilité est la même pour tous les utilisateurs du réseau.

L'exemple développé illustre les idées principales utilisées dans la modélisation des grands réseaux en utilisant la géométrie stochastique. Dans ce qui suit, nous présentons les principales contributions de ce travail. Une description plus détaillée des définitions et résultats sur les processus ponctuels peuvent être trouvées dans le Chapitre 4.

1.5 Modèles de réseaux coopératifs

L'idée principale des réseaux sans fils coopératifs [11, 12] est d' avoir deux ou plusieurs utilisateurs partageant leurs antennes pour atteindre un taux de transfert ou fiabilité supérieure. En général, l' analyse de ces réseaux considèrent seulement un canal isolé et la présence de bruit additive décorrélé aux récepteurs. Dans ce contexte, les avantages prédits par l'utilisation de la coopération sont grandes ; cependant, les canaux sans fils ne sont pas toujours

isolés et c'est intéressant de considérer l'interaction entre les nœuds voisins. De plus, si les nœuds coopèrent pour transmettre les messages d'autres utilisateurs sans coordination, le niveaux d'interférence pourrait réduire les effets positifs de la coopération au point de la rendre inutile. L'objectif de cette thèse est d'étudier certains aspects des réseaux coopératifs sans fils en considérant les effets adverses de l'interaction entre les utilisateurs. Dans les sections suivantes nous présentons un résumé des contributions principales de cette thèse.

1.5.1 Utilisation de relais dans les réseaux sans fils

Le canal coopératif plus simple et intuitif d'imaginer est, peut-être, le canal de relai (*relay channel*) [13], dans lequel un nœud appelé relai (*relay*) aide un émetteur à transmettre un message à un nœud destination. Une représentation du canal de relai gaussien est présentée dans la Figure 1.4 et les équations que le représentent sont :

$$Y_{r,i} = g_{sr}X_{s,i} + Z_{r,i} \tag{1.25}$$

$$Y_{d,i} = g_{sd}X_{s,i} + g_{rd}X_{r,i} + Z_{d,i}, \qquad i = 1, 2, \ldots, \tag{1.26}$$

où les bruits $\{Z_{r,i}\}_i$ et $\{Z_{d,i}\}_i$ sont des séquences de bruit blanc gaussien complexe indépendantes entre eux et des émissions des utilisateurs, et avec variances peut-être différentes. $X_{s,i}$ et $X_{r,i}$ sont les symboles émis par la source et le relai à l'instant i. Les techniques de codage

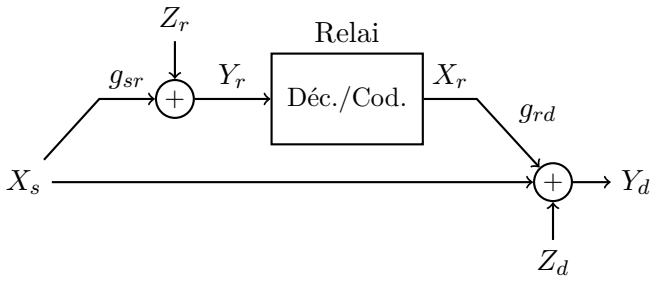

FIGURE 1.4: Modèle mathématique du canal de relai.

principales pour le canal de relai ont été introduites en [14] : décodez-et-renvoyez (*decode-and-forward*, DF), et comprimez-et-renvoyez (*compress-and-forward* ,CF). Dans la première, le relai décode le message de la source, il le codifie à nouveau, et le renvoie à la destination. La destination essaie de décoder le message de la source en utilisant les deux transmissions. Pour CF, le relai ne décode pas le message de la source, mais il sélectionne d'une librairie, connue par le relai et la destination, un nouveau message qui représente la transmission de la source. Alors, il transmet l'index de ce message à la destination, qui l'utilise pour décoder le message avec la transmission de la source. Les avantages d'utiliser ces protocoles quand on

considère un canal de relai isolé avec du bruit décorrelationé, sont importantes [15, 16]. Dans les chapitres 5 et 6 ces résultats sont étudiés dans le contexte d'un grand réseaux sans fils, où l'interférence entre les utilisateurs est présente.

Dans le Chapitre 5 nous étudions la performance, évaluée par la probabilité d'outage, atteignable dans un canal de relai full-duplex quand les nœuds opèrent dans un grand réseau sans fils où les émetteurs interférants sont modelés avec un processus ponctuel de Poisson homogène $\Phi = \{x_i\}$ de densité λ. Cet analyse diffère notamment de l' analyse traditionnelle où on considère du bruit blanc et decorrélé aux récepteurs (relai et destination). Pour chaque réalisation du processus ponctuel, les signaux temporels d'interférence sont corrélés, car ils sont générés par les mêmes émetteurs. De plus, cette corrélation est aléatoire si nous considérons que c'est une fonction du processus. Un protocole est *full-duplex* quand le relai peut transmettre et recevoir au même temps et *half-duplex* quand le relai écoute l'émission de la source pendant une première partie du temps, et après il transmet pour le reste du temps.

Nous considérons que le nœud source est à l'origine de coordonnées spatiale, le relai est à la position r et la destination à d. Avec la notation de l'équation (1.16) les signaux reçus à la destination et au relai sont (sans indiquer l'index du temps) :

$$Y_r = h_{sr}\sqrt{l_{sr}}X_s + \underbrace{\sum_{x\in\Phi} h_{xr} l(x,r)^{\frac{1}{2}} X_x}_{\tilde{Z}_r}$$

$$Y_d = h_{sd}\sqrt{l_{sd}}X_s + h_{rd}\sqrt{l_{rd}}X_r + \underbrace{\sum_{x\in\Phi} h_{xd} l(x,d)^{\frac{1}{2}} X_x}_{\tilde{Z}_d},$$

où $\{X_{x_i}\}_i$ sont les symboles émis par les autres émetteurs du réseau. Si les codes employés par les transmetteurs sont gaussiens, de moyenne nulle et de variance l' unité, et la destination traite l'interférence comme bruit pour le décodage, alors pour chaque réalisation du processus, les signaux temporales d'interférence $\tilde{Z}_r$ et $\tilde{Z}_d$, sont des variables aléatoires gaussiennes complexes, circulairement symétriques, de moyenne nule et variances :

$$I_r \triangleq \mathbb{E}\left[|\tilde{Z}_r|^2|\tilde{\Phi}\right] = \sum_{x\in\Phi} |h_{xr}|^2 l(x,r), \tag{1.27}$$

$$I_d \triangleq \mathbb{E}\left[|\tilde{Z}_d|^2|\tilde{\Phi}\right] = \sum_{x\in\Phi} |h_{xd}|^2 l(x,d). \tag{1.28}$$

Les signaux temporelles sont aussi corrélés dans chaque instant, et son coefficient de corrélation est :

$$\rho_N = \frac{\mathbb{E}[\tilde{Z}_r \tilde{Z}_d^*|\tilde{\Phi}]}{\sqrt{I_r I_d}}. \tag{1.29}$$

Comme les codes sont gaussiens, il est possible de trouver les taux de transfert atteignables de DF et CF et déterminer leurs événements d'outage. Dans le cas de DF, si le relai ne décode pas

l'émission de la source, la destination peut également essayer de décoder le message comme si le relai ne fût pas présent. Aussi, le taux de transmission atteignable avec DF ne dépend pas explicitement de la corrélation (1.29) parce que le relai est obligé à décoder le message de la source. Pour cette raison, pour DF il est possible de trouver une expression pour la probabilité d'outage (voir la Section 5.3.1 pour les détails) :

Théorème 1.1 (OP de DF)**.** *La probabilité d'*outage *de DF est :*

$$P_{out,DF}(R,\rho) = 1 - \mathbb{P}\left(\mathcal{A}_{DF}^{c}(R,\rho) \cap \mathcal{B}_{DF}^{c}(R,\rho)\right) - \mathbb{P}\left(\mathcal{A}_{DF}(R,\rho) \cap \mathcal{A}_{DT}^{c}\right), \tag{1.30}$$

où $\mathcal{A}_{DF}$ représente l'événement que le relai ne décode pas l'émission de la source, $\mathcal{B}_{DF}$ indique que la destination ne décode pas le message en utilisant les deux transmissions du relai et de la source, et $\mathcal{A}_{DT}$, que la destination ne décode pas la transmission de la source quand le relai ne réussit pas à décoder. R est le taux de transfert auquel la source essaie de transmettre à la destination, et ρ est le coefficient de corrélation entre les symboles émis par la source et le relai.

Les termes de (1.30), quand $||r-d|| \neq ||d||$ où $\rho \neq 0$ peuvent être écrits comme suit :

$$\mathbb{P}\left(\mathcal{A}_{DF}^{c}(R,\rho) \cap \mathcal{B}_{DF}^{c}(R,\rho)\right) = \frac{\mu_2 \mathcal{L}_{I_d,I_r}\left(\frac{T(R)}{\mu_2}, \frac{T(R)}{\mu_3}\right) - \mu_1 \mathcal{L}_{I_d,I_r}\left(\frac{T(R)}{\mu_1}, \frac{T(R)}{\mu_3}\right)}{\mu_2 - \mu_1}, \tag{1.31}$$

où :

$$\mu_1 = \frac{1}{2}\left[l_{sd} + l_{rd} - \left((l_{sd}-l_{rd})^2 + 4 l_{sd} l_{rd} |\rho|^2\right)^{\frac{1}{2}}\right], \tag{1.32}$$

$$\mu_2 = \frac{1}{2}\left[(l_{sd} + l_{rd}) + \left((l_{sd}-l_{rd})^2 + 4 l_{sd} l_{rd} |\rho|^2\right)^{\frac{1}{2}}\right], \tag{1.33}$$

$$\mu_3 = l_{sr}\left(1 - |\rho|^2\right), \tag{1.34}$$

et $\mathcal{L}_{I_d,I_r}(\omega_1,\omega_2)$ est la transformée de Laplace conjointe des interférences, donnée par (5.5), et $T = T(R) = 2^R - 1$. De plus, quand $||r-d|| = D$ et $\rho = 0$, alors $\mu_1 = \mu_2$ et :

$$\mathbb{P}\left(\mathcal{A}_{DF}^{c}(R,\rho) \cap \mathcal{B}_{DF}^{c}(R,\rho)\right) = \mathcal{L}_{I_d,I_r}\left(\frac{T(R)}{\mu_1}, \frac{T(R)}{\mu_3}\right) - \frac{T}{\mu_1} \frac{d\mathcal{L}_{I_d,I_r}(\omega_1, T(R)/\mu_3)}{d\omega_1}\bigg|_{\omega_1 = T(R)/\mu_1}.$$

Finalement :

$$\mathbb{P}\left(\mathcal{A}_{DF} \cap \mathcal{A}_{DT}^{c}\right) = \mathcal{L}_{I_d}\left(\frac{T(R)}{l_{sd}}\right) - \mathcal{L}_{I_d,I_r}\left(\frac{T(R)}{l_{sd}}, \frac{T(R)}{\mu_3}\right). \tag{1.35}$$

Il est possible de prouver (Lemma 5.3) que, dans le régime typique d'opération des réseaux sans fils, le coefficient de corrélation ρ qui minimise la probabilité d'*outage* et $\rho = 0$, et avec ça, simplifie les expressions du théorème précédent.

L'analyse de CF est plus compliquée que celle du DF, car en CF le relai comprime l'information reçue de la source sans décoder le message, et ça implique que la corrélation entre

les interférences (1.29) apparaissent explicitement dans le taux atteignable. C'est pour ça que l'événement d'*outage* de CF, noté $\mathcal{O}_{CF}(R, n_c, \rho_N)$, va dépendre du taux R et de la corrélation ρ_N. En plus, il y a aussi un paramètre $n_c > 0$ qui contrôle le niveau de compression du message au relai. La distribution de la corrélation ρ_N dépend du processus, et donc, l'expression exacte de la probabilité d'outage ne peut pas être trouvée. Pour cette raison, nous introduisons le lemme suivant (Section 5.3.2) :

Lemme 1.1. *La taux atteignable par CF avec une corrélation arbitraire ρ_N entre les bruits aux recepteurs est au plus un bit pire que le taux obtenu avec les bruits décorrélés :*

$$R_{CF}(\rho_N, n_c) \geq R_{CF}(0, n_c) - 1. \tag{1.36}$$

Avec ce résultat nous pouvons trouver un majorant de la probabilité d'outage :

Théorème 1.2. *Pour un taux essayé R la probabilité d'outage de CF peut être majorée comme suit :*

$$\begin{aligned} P_{out,CF}(R, n_c) &\leq \mathbb{P}(\mathcal{O}_{CF}(R+1, n_c, 0)) \\ &= \mathbb{P}(\mathcal{A}_{CF}(R+1, 0) \cup \mathcal{B}_{CF}(n_c, 0)). \end{aligned} \tag{1.37}$$

L'événement $\mathcal{A}_{CF}$ indique que la destination ne peut pas décoder la transmission du relai (la valeur de n_c choisie est trop haute), et $\mathcal{B}_{CF}$ indique la destination ne peut pas décoder le message avec les deux transmissions de la source et le relai.

La probabilité d'outage avec $\rho_N = 0$ peut être majorée comme suit :

$$\begin{aligned} \mathbb{P}(\mathcal{A}_{CF}(R,0)) \leq 1 - e^{-\frac{T(R)n_c}{l_{sr}}} \mathcal{L}_{I_r}\left(\frac{T(R)}{l_{sr}}\right) - \Bigg[\sum_{\tilde{n}=0}^{\tilde{N}-1} e^{\frac{\tilde{n}n_cT(R)}{\tilde{N}l_{sr}}} \mathcal{L}_{I_d,I_r}\left(\frac{(\tilde{N}-\tilde{n})T(R)}{\tilde{N}l_{sd}}, \frac{\tilde{n}T(R)}{\tilde{N}l_{sr}}\right) \\ - e^{\frac{(\tilde{n}+1)n_cT(R)}{\tilde{N}l_{sr}}} \mathcal{L}_{I_d,I_r}\left(\frac{(\tilde{N}-\tilde{n})T(R)}{\tilde{N}l_{sd}}, \frac{(\tilde{n}+1)T(R)}{\tilde{N}l_{sr}}\right)\Bigg], \end{aligned} \tag{1.38}$$

avec $\tilde{N} \in \mathbb{N}$, et :

$$\begin{aligned} &\mathbb{P}(\bar{\mathcal{A}}_{CF}(R,0) \cap \mathcal{B}_{CF}(n_c,0)) \leq \\ &\qquad 1 - \mathbb{E}\left[\mathcal{L}_{I_d,I_r}\left(\frac{(1+T(R))l_{sr}|h_{sr}|^2}{T(R)n_cl_{rd}}, \frac{(1+T(R))l_{sd}|h_{sd}|^2}{T(R)n_cl_{rd}}\right)\right]. \end{aligned} \tag{1.39}$$

L'espérance est de h_{sr} et h_{sd}, avec $T(R) = 2^R - 1$.

Après d'avoir étudié les deux protocoles mentionnés, nous faisons une comparaison entre les probabilités d'outage de ces protocoles et la probabilité d'outage d'une transmission point à point et d'un protocole DF half-duplex, pour lequel nous trouvons un minorant de la probabilité d'outage (Theorème 5.5). Dans la Section 5.4 nous présentons les résultats numériques

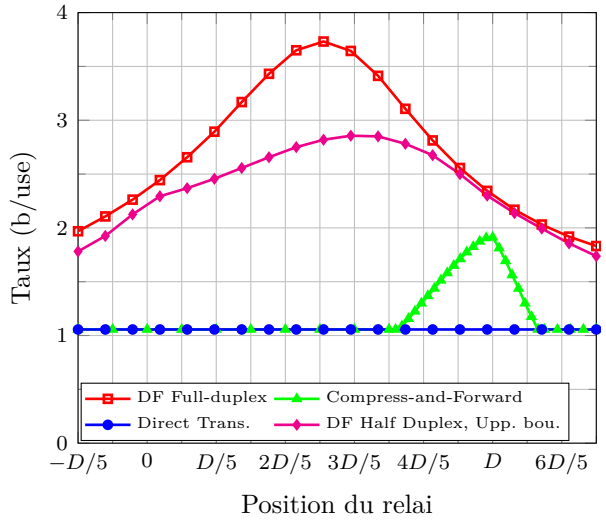

FIGURE 1.5: Taux maximale atteignable avec chaque protocole quand le relai est entre la source et la destination et une contrainte de OP maximale de 0.05 est demandée. La OP de full-duplex DF vient du Théorème 5.2, half-duplex DF du Théorème 5.5, CF du Théorème 5.7 et transmission directe de (5.40). $d = (10, 0)$, $\alpha = 4$.

pour comparer les protocoles. Spécialement nous faisons mention de la Figure 1.5, où nous étudions le taux maximal qu' on peut atteindre avec chaque protocole quand le relai est aligné avec la source et la destination, et il y a une restriction de la probabilité d'outage maximale de 0.05. Les paramètres principaux sont dans la figure. Nous pouvons observer que DF full-duplex a une meilleure performance, spécialement quand le relai est à la même distance de la source et de la destination. Dans le cas du protocole *half-duplex*, la courbe est obtenue avec un minorant de la probabilité d'*outage* et ça donne un majorant du taux maximal. Aussi, CF ne réussit pas à être meilleur que DF, et ça indique que le majorant de la probabilité d'*outage* est possiblement un peu prudente, c'est-à-dire, que l'effet de la corrélation entre les bruits est importante. Pour vérifier ça, dans la Figure 5.8 nous représentons les régions spatiales où CF, DF *full-duplex*, ou une transmission point à point ont la meilleur probabilité d'*outage*. La probabilité d'*outage* de CF est obtenue avec une simulation Monte Carlo. Nous pouvons voir qu' effectivement autour de la destination, CF est meilleure que DF, mais DF est supérieure dans le reste de la région représentée. Notons aussi que pour CF, il est nécessaire de déterminer la valeur de n_c qui minimise la probabilité d'outage.

Dans le Chapitre 6 nous étudions une situation plus générale dans laquelle les émetteurs qui causent l'interférence peuvent aussi utiliser un relai. En particulier, nous considérons un réseau décentralisé dans lequelle chaque émetteur peut utiliser un relai, qui est le nœud voisin

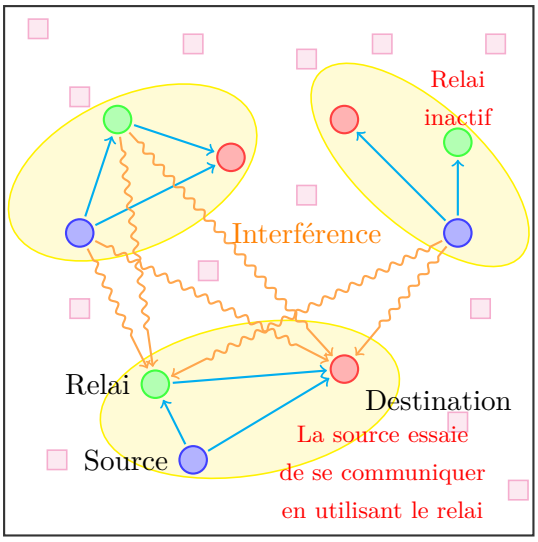

FIGURE 1.6: Le réseau a des groupes d'utilisateurs (en jaune) qui utilisent DF ou transmission point à point. Les relais (en vert) sont choisis de l'ensemble des nœuds inactifs du réseau (carrés en magenta) pour aider les sources (en bleu) à transmettre vers leurs destinations (en rouge).

plus proche qui soit inactive (voir Figure 1.6). Nous introduisons une stratégie d'activation des relais dans laquelle ils sont activés de façon indépendant de tout. Les expressions des interférences sont dans ce cas beaucoup plus compliquées car il va y avoir des groupes de nœuds avec deux émetteurs (une source et son relai). Également, nous pouvons trouver une expression pour la probabilité d'*outage* du protocole (Section 6.3) :

Théorème 1.3. *La probabilité d'*outage *des émetteurs du réseau* $P_{out,mix}$ *est :*

$$P_{out,mix}(R) = \mathbb{P}(\varepsilon_0 = 0)\left[1 - \mathcal{L}_{I_d}\left(\frac{T(R)}{l_{sd}}\right)\right] + \mathbb{P}(\varepsilon_0 = 1) \times \mathbb{E}_r\left[\frac{D^\alpha \mathcal{L}_{I_d,I_r}\left(\frac{T(R)}{l_{rd}}, \frac{T(R)}{l_{sr}}\right) - ||r-d||^\alpha \mathcal{L}_{I_d,I_r}\left(\frac{T(R)}{l_{sd}}, \frac{T(R)}{l_{sr}}\right)}{D^\alpha - ||r-d||^\alpha}\right], \quad (1.40)$$

où, comme dans les cas précédents, $\mathcal{L}_{I_d,I_r}(\cdot,\cdot)$ *est la transformée de Laplace des interférences, donnée par (5.5). La variable aléatoire Bernoulli* ε_0 *indique si la source à l'origine des coordonnées a un relai.*

Avec la probabilité d'*outage* d'un utilisateur, nous étudions la valeur de la probabilité d'activation des relais p_r qui minimise cette probabilité. Pour ça nous considerons que les relais son près de la source et que le réseau se trouve dans son régime d'opération normale où la probabilité d'*outage* est petite. Dans ce régime, nous prouvons (Théorème 6.3) que la

stratégie qui minimise la probabilité d'*outage* est d'activer ou éteindre tous les relais au même temps, dépendent des paramètres du réseau comme la distance à la destination, la densité des émetteurs, le taux, etc. Cependant, ce résultat n'est pas valide quand la probabilité d'*outage* est grande, c'est-à-dire, quand le réseau ne se trouve pas dans son régime habituel.

1.5.2 Communications entre dispositifs mobiles (D2D) aux réseaux cellulaires

Le trafic dans les réseaux cellulaires a augmenté fortement ces dernières années, principalement en raison de la transition des systèmes de communication purement de voix (2G et précédents) aux systèmes avec prédominance du trafic IP (à partir de 3G). Pour cette raison, il est nécessaire d'avoir de nouvelles techniques pour satisfaire les nouvelles exigences, en considérant une meilleure utilisation du spectre radioélectrique, des meilleurs systèmes de modulation, réduction des distances de transmission, etc.. La stratégie qui augmente plus fortement la performance des réseaux est la réduction de la taille des cellules et des distances de transmission [17], conduisant à une plus grande réutilisation du spectre et à une plus grande efficacitée spectrale. Cette observation introduira des modifications importantes dans l'infrastructure des réseaux cellulaires, et changera le paradigme homogène des réseaux cellulaires, avec des stations de base dispersées de longue portée. Le nouveau paradigme va être hétérogène, avec un réseau de stations de base très dense avec des tailles de cellule diverses, et avec des nouveaux types d'interaction avec les bases et entre les téléphones mobiles. Une des stratégies possibles, qui seront utilisées dans le contexte des réseaux mobiles pour atteindre cet objectif, est celui des communications mobiles (*device-to-device*, D2D) [18], où deux utilisateurs mobiles qui sont proches peuvent être connectés directement via une transmission de courte distance, évitant une connexion avec la station de base [19]. Cette stratégie contribuera, par exemple, à la réutilisation des fréquences, à l'efficacité énergétique et à la réduction des temps de latence. Selon les estimations actuelles [1] une des causes principales de l'augmentation du trafic pour les prochaines années seront les services de transmission de vidéo sur demande (*on-demand)*, où un grand nombre d'utilisateurs accèdent à une petite bibliothèque de vidéos, de façon asynchronique et donc une transmission unique ne peut pas être utilisée par tous les utilisateurs [20]. Dans ce contexte, le réseau pourrait profiter de la mémoire disponible dans les dispositifs mobiles, et constituer un système de stockage distribué pour l'échange de vidéos entre les utilisateurs, au lieu d'utiliser la station de base. En outre, si les échanges ont lieu en dehors des bandes de fréquence cellulaires (*out-of-band D2D*), c'est possible aussi d'augmenter la réutilisation des fréquences.

Le but du Chapitre 7 est d'étudier certains aspects de la performance des stratégies de partage de vidéos par *out-of-band* D2D, à l'aide d'un modèle de processus ponctuel. En

particulier, on suppose que dans un certain moment et dans une certaine zone du réseau il y a de nombreux utilisateurs qui ont des vidéos stockés, et d'autres utilisateurs qui demandent des vidéos pour les regarder. Nous souhaitons étudier la fraction des demandes pour les vidéos qui peuvent être satisfaites par D2D, c'est-à-dire, par le biais des émissions locales, plutôt que par la station de base. Cette fraction est un indicateur des avantages économiques potentiels et en termes de qualité de service que peut être obtenue en D2D. Ce problème doit être étudié ayant en considération les aspects suivants :

- La stratégie de stockage dans les utilisateurs.
- Le problème de comment lier les utilisateurs pour faire l'échange (qui va transmettre le vidéo demandé).
- Les problèmes associés à la transmission et coordination entre les utilisateurs.

Pour étudier ce problème, nous introduisons un modèle de processus ponctuel pour donner un cadre théorique pour l'analyse de stratégies D2D. Ce modèle a les caractéristiques suivantes :

- Il considère l'existence de nœuds qui demandent vidéos (destinations) et de nœuds coopératifs (avec vidéos stockés), qui son distribués dans l'espace comme deux processus ponctuels de Poisson indépendants, Φ_r et Φ_u respectivement. Les canaux de communication sont modelés comme canaux *slow-fading*.
- Les utilisateurs son regroupés dans l'espace en groupes ou *clusters* dans lesquels des vidéos sont échangés. Les *clusters* sont des disques disjoints de rayon R_c. Les centres des *clusters* forment un processus ponctuel Φ_p de noyau dure (*hard core*) qui garantisse une distance minimale $\delta \geq 2R_c$ entre les centres, et que les groupes soient disjoints. Cette idée permets de construire un processus de nœuds regroupés selon la définition suivante :

  Définition 1.5.1. *Pour un réseau avec des* clusters *de rayon $R_c > 0$, le processus d'utilisateurs regroupés Φ_c est obtenu à partir de Φ_u et Φ_r comme suit :*

  $$\Phi_c = \bigcup_{x \in \Phi_p} \mathcal{B}(x, R_c) \cap (\Phi_r \cup \Phi_u), \tag{1.41}$$

  où $\Phi_p = \{x_i\}$ est un processus ponctuel hard-core *de densité $\lambda_p > 0$ et distance minimale entre points $\delta \geq 2R_c$.*

- Pour l'échange de vidéos dans les groupes, nous introduisons une famille de protocoles admissibles, qui sont utilisés par les nœuds de chaque groupe. Un protocole admissible est formé par un couple $(\Phi_p, \mathcal{F})$, où Φ_p est un processus donnant les centres des groupes et $\mathcal{F}$ est un stratégie d'échange de vidéos dans un groupe.

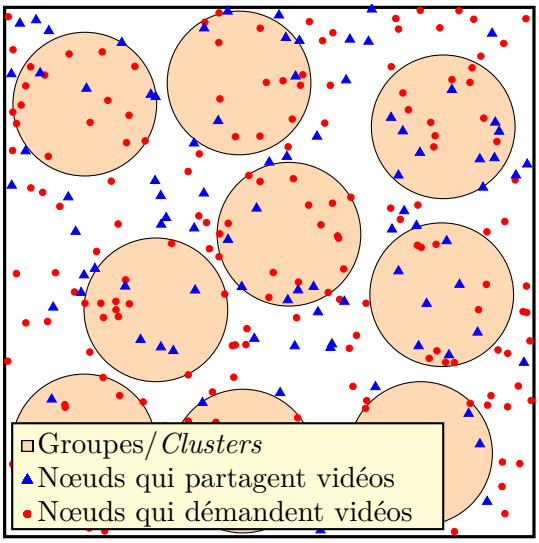

FIGURE 1.7: Représentation du réseau avec groupes. Les nœuds dans le groupes forment le processus Φ_c selon (1.41), où les échanges de vidéos ont lieu.

Dans ce contexte, nous introduisons deux métriques qui caractérisent la fraction de demandes que peuvent être satisfaites par D2D :

- Une *métrique globale*, T_G, définie comme le quotient entre la densité de demandes satisfaites par D2D et la densité total de demandes du réseau.
- Une *métrique locale*, T_L, définie comme le quotient entre le nombre moyen de demandes satisfaites dans un groupe et le nombre moyen de demandes dans un groupe du réseau.

Ces métriques sont définies afin d'étudier la quantité de demandes qui peuvent être satisfaites par D2D et elles considèrent les trois aspects du problème mentionné, nous devons aussi considérer des exigences de qualité de services qui pourraient être nécessaires pour l'échange de vidéos. Par exemple, un grand nombre de demandes pourraient être satisfaites mais avec un taux de transmission très bas. Pour cette raison, nous introduisons trois régions de compromis d'intérêt qui considèrent les deux métriques définies mais aussi des exigences de taux moyen pour les transmissions :

Définition 1.5.2 (Régions de compromis)**.** *Nous considérons les régions de compromis :*

- Région de compromis de la métrique globale : *un couple* (r, t) *de taux moyen et métrique globale est dit atteignable s'il existe un protocole admissible* $(\Phi_p^\dagger, \mathcal{F}^\dagger)$ *avec un taux de*

transmission R et densité λ_p qui satisfait :

$$\begin{cases} T_G(\mathcal{F}^\dagger, R) \geq t, \\ \bar{R}(\mathcal{F}^\dagger, R) \geq r. \end{cases} \tag{1.42}$$

L'ensemble de tous les couples atteignables (r, t) est la région de compromis de la métrique global.

- Région de compromis de la métrique locale : *un triplet (r, t, λ_l) de taux moyen, métrique locale et densité de groupes est dit atteignable s'il existe un protocole admissible $(\Phi_p^\dagger, \mathcal{F}^\dagger)$ qui satisfait :*

$$\begin{cases} T_L(\mathcal{F}^\dagger, R) \geq t, \\ \bar{R}(\mathcal{F}^\dagger, R) \geq r, \\ \lambda_p(R_c, \delta) \geq \lambda_l. \end{cases} \tag{1.43}$$

L'ensemble de tous les triplets atteignables (r, t, λ_l) est la région de compromis de la métrique locale.

- Région de compromis des métriques globale et locale : *pour un taux essayé R, un couple (t_g, t_c) de métrique globale et locale est dit atteignable s'il existe un protocole admissible $(\Phi_p^\dagger, \mathcal{F}^\dagger)$ de taux R et densité de groupes λ_p qui satisfait :*

$$\begin{cases} T_G(\mathcal{F}^\dagger, R) \geq t_g, \\ T_L(\mathcal{F}^\dagger, R) \geq t_c. \end{cases} \tag{1.44}$$

L'ensemble de tous les couples atteignables (t_g, t_c) est la région de compromis de compromis des métriques globale et locale.

La détermination des régions de compromis implique trouver le ou les protocoles optimales, et ça c'est un problème très compliqué. Pour cette raison nous analysons un protocole spécifique qui donne une borne intérieure des régions. Dans la stratégie proposée, les nœuds sont reliés en couples et les vidéos sont échangés par le biais de transmission point à point. Dans chaque groupe, l'interférence est réduite avec une stratégie d'accès multiple à répartition dans le temps (TDMA), qui divise le canal en intervalles de temps, chacun occupé par une transmission. L'analyse de la stratégie permet d'observer que l'interférence pendant chaque transmission n'est pas constante, et ça rend nécessaire une analyse spéciale d'interférence.

Capítulo 2

Concepción y Análisis de Redes Inalámbricas de Grandes Dimensiones Utilizando Técnicas de Geometría Estocástica

2.1 Preliminares

Las comunicaciones inalámbricas son el segmento de mayor crecimiento en la industria de las telecomunicaciones. El surgimiento de nuevas aplicaciones es constante, convergiendo a una visión en la que cada dispositivo electrónico en la Tierra estará conectado a una red. Dichas aplicaciones cubren los campos más diversos como comunicaciones entre máquinas (*machine-to-machine*), redes eléctricas inteligentes (*smart-grids*), redes de sensores inalámbricos (*wireless sensor networks*), redes de seguridad pública, etc. Uno de los segmentos que ha experimentado un crecimiento más vertiginoso es el de las comunicaciones celulares. Dicho crecimiento ha sido principalmente causado por un cambio en el tipo de tráfico que transportan las redes, que ha pasado de ser principalmente de voz, a tráfico de datos IP, alimentado por el advenimiento de los teléfonos inteligentes, tabletas y *streaming* de video [1]. Los nuevos requisitos representan desafíos importantes para la ingeniería de las comunicaciones, requiriendo novedosos y sustanciales avances en el campo. En este contexto, es necesario definir nuevos paradigmas y arquitecturas de red, que redefinan las interacciones entre los nodos, permitiendo maximizar la eficiencia espectral y la utilización de recursos. La cooperación entre los nodos, por lo tanto, jugará un rol importante en esta tarea y sus aspectos principales deben ser estudiados mediante modelos matemáticos que sinteticen los elementos principales

que caracterizan a las redes inalámbricas.

La geometría estocástica [2] y, en particular, la teoría de procesos puntuales son herramientas matemáticas especialmente adaptadas para el modelado de los aspectos principales de redes inalámbricas, donde los nodos están distribuidos en forma aleatoria en el espacio. El objetivo global del presente trabajo, que se desarrollará más en detalle en este capítulo, es el estudio de aspectos cooperativos de redes inalámbricas, utilizando las herramientas de los procesos puntuales. En lo que sigue, se resumirán los aspectos fundamentales del trabajo realizado; en cada caso se indicarán las referencias pertinentes a la parte principal en inglés, donde los temas se desarrollan con más detalle.

2.2 Modelos de canales inalámbricos empleados

El medio inalámbrico es, en general, un medio más adverso para establecer una comunicación que los medios cableados tradicionales. Las transmisiones inalámbricas son casi siempre omnidireccionales, de modo que, además de la propagación por la distancia recorrida, existen otros fenómenos de atenuación y distorsión vinculados con la geometría del ambiente donde se propagan las ondas. Fundamentalmente pueden mencionarse tres fenómenos [3]:

- La atenuación que experimenta una onda electromagnética que se propaga en función de la distancia recorrida, conocida como pérdida por camino (*path loss*).
- Desvanecimiento de gran escala o *shadow fading*, que se refiere a la atenuación causada por objetos de grandes dimensiones que se encuentran en el medio de propagación.
- Desvanecimiento de pequeña escala o por multicamino (*multipath fading* o *fading* a secas), que ocurre cuando en la misma señal se propaga por un gran número de caminos distintos de distinta longitud (distintos retardos), que se superponen en el destino.

Dado que los fenómenos de desvanecimiento no son predecibles en general, porque dependen de la geometría del medio y pueden cambiar en el tiempo, habitualmente se recurre a modelos estadísticos de dichos fenómenos. Por otro lado, para la atenuación por distancia se consideran modelos determinísticos que son función de la distancia entre el transmisor y el receptor, y del medio de propagación.

En general, los canales inalámbricos son modelados como canales lineales variantes en el tiempo, caracterizados por una respuesta impulsiva variante $c(t, \xi)$, que representa la respuesta del canal en el instante t a un impulso que ingresó al canal ξ unidades de tiempo antes. Dado que las variaciones de dicha respuesta son en general impredecibles, es habitual caracterizarla por medio de un proceso estocástico. Estudiando la autocorrelación de dicho proceso, y bajo

ciertas hipótesis razonables, pueden definirse dos parámetros que caracterizan, a grandes rasgos, los comportamientos que puede tener un canal inalámbrico. Dichos parámetros son:

- La dispersión por multicaminos (*multipath spread*) T_m, que da un estimación del tiempo que demora la salida del canal en descorrelacionarse de su entrada. El recíproco de dicho parámetro es el llamado ancho de banda de coherencia (*coherence bandwidth*) B_{coh}. Un canal pasabanda de ancho de banda W será no selectivo en frecuencia (*flat fading* o *frequency non-selective*) cuando satisfaga la condición $W \ll B_{\text{coh}}$. En ese caso, todas las frecuencias a la entrada del canal se verán afectas de la misma manera.

- La dispersión Doppler del canal (*Doppler spread*) B_d que indica qué tan rápido cambia el canal en el tiempo. El recíproco de dicho parámetro es el llamado tiempo de coherencia del canal (*coherence time*), T_{coh}. Si la duración de un bloque T_s satisface $T_s \ll T_{\text{coh}}$ se dice que el canal es de desvanecimiento lento (*slow fading*).

En este trabajo nos enfocaremos en los canales no selectivos y de desvanecimiento lento, lo que implica que el equivalente en banda base del canal de comunicaciones puede escribirse como:

$$y(t) = gx(t), \tag{2.1}$$

donde $x(t)$ es la señal transmitida, $y(t)$ la señal recibida y g es la ganancia compleja del canal (constante durante la transmisión por la hipótesis de desvanecimiento lento). Dicha ganancia es seleccionada de acuerdo a un modelo que considere los fenómenos de atenuación que afecten al canal inalámbrico que se mencionaron anteriormente. En general no se consideran todos los fenómenos al mismo tiempo. Un modelo estándar considera solamente la presencia de *path loss* y de *fading*, ignorando el *shadow fading*. En ese caso, para un transmisor ubicado en u y un receptor ubicado en v, la ganancia del canal puede escribirse como:

$$g = l(u, v)^{\frac{1}{2}} h, \tag{2.2}$$

donde $l(u, v)$ es una función que representa el *path loss* entre u y v, y h es una constante compleja que contiene el *fading*. Normalmente dicha función depende sólo de la distancia entre los puntos, es decir, $l(u, v) \equiv l(||u - v||)$. Existen varios ejemplos de esas funciones; la más habitual, que consideraremos en este trabajo, es $l(u, v) = ||u - v||^{-\frac{\alpha}{2}}$, donde α es el exponente de atenuación con la distancia, normalmente en el intervalo $(2,\ 6)$. Existen diversos modelos estadísticos para seleccionar el coeficiente complejo h. El más habitual, que utilizaremos en general en este trabajo, es el llamado *fading* Rayleigh, en el cual se asume que $|h|$ es una variable Rayleigh y la fase de h es uniforme en $[0, 2\pi)$. Esto implica que $|h|^2$ seguirá una distribución exponencial.

Los canales que consideraremos en este trabajo, como es habitual en la bibliografía, son a tiempo discreto, pudiendo en general ser convertidos a canales en tiempo continuo sin pérdida, mediante un procedimiento de muestreo adecuado (ver [4]). Por lo tanto, resumiendo los resultados presentados en esta sección, el modelo de desvanecimiento lento no selectivo que se empleará en este trabajo será el siguiente:

$$y_i = l(u,v)^{-\frac{\alpha}{2}} h x_i, \tag{2.3}$$

donde $i = 1, 2, \ldots$ denota el tiempo discreto. Algunos detalles adicionales sobre la derivación de estos modelos pueden hallarse en el Apéndice A.

2.3 Medidas de capacidad para canales de desvanecimiento lento

Uno de los objetivos fundamentales en el estudio de los sistemas de comunicaciones es el de conocer las tasas máximas que pueden ser transmitidas en forma confiable a través de un canal de comunicaciones. La primera formulación rigurosa de dicho problema y los primeros resultados fueron desarrollados por C. E. Shannon en 1948 [5]. Sus trabajos resultaron tan fundamentales que dieron lugar a la llamada Teoría de la Información, que posee hoy en día infinidad de aplicaciones en los campos más diversos [4].

Los primeros resultados de Shannon se orientaron al estudio de los canales punto a punto, es decir, entre sólo un transmisor y un receptor, cuyo diagrama puede verse en la Figura 2.1. Los elementos principales que constituyen el canal son los siguientes:

- Un alfabeto de símbolos $\mathcal{X}$ que pueden transmitirse por el canal.
- Un alfabeto de símbolos $\mathcal{Y}$ que pueden ser observados a la salida del canal.
- Una distribución de probabilidad condicional $p_{Y^n|X^n}$ que, dada un secuencia $X^n \in \mathcal{X}^n$ de largo n a la entrada del canal, se utiliza para generar una secuencia $Y^n \in \mathcal{Y}^n$ de largo n a la salida del canal.

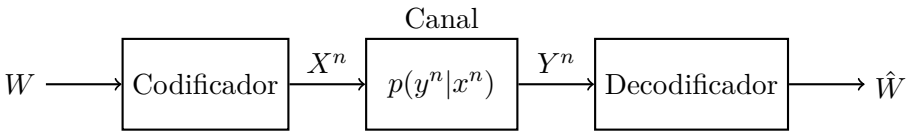

Figura 2.1: Diagrama en bloques de un canal de comunicaciones punto a punto.

Para realizar una transmisión de largo n en el canal, se utilizará un código (M, n), constituido de la siguiente manera:

- Un conjunto de mensajes $\{1, \ldots, M\}$ de los cuales se elige uno para ser transmitido, denotado como W.
- En el transmisor: una función de codificación $X^n : \{1, \ldots, M\} \to \mathcal{X}^n$, que asigna a cada mensaje una secuencia de largo n, llamada palabra de código, para ser transmitida por el canal.
- En el receptor: una función de decodificación $Y^n : \mathcal{Y}^n \to \{1, \ldots, M\}$ que genera un estimado del mensaje transmitido, denotado como $\hat{W}$.

La tasa R de un código (M, n) (en bits/uso de canal) se define como:

$$R = \frac{\log_2(M)}{n}. \tag{2.4}$$

La probabilidad de error de una palabra de código se define como:

$$P_{e,i} = \mathbb{P}(\hat{W} \neq W | W = i). \tag{2.5}$$

Una tasa R se dice alcanzable (*achievable*) si existe una secuencia de códigos $(2^{nR}, n)$, tales que :

$$\max_{i \in \{1, \ldots, 2^{nR}\}} P_{e,i} \underset{n \to \infty}{\longrightarrow} 0, \tag{2.6}$$

es decir, la máxima probabilidad de error tiende a cero cuando se utiliza un código más largo. La capacidad del canal se define entonces como el supremo de todas las tasas alcanzables, es decir:

$$C = \sup\{R : R \text{ es alcanzable}\}. \tag{2.7}$$

Shannon fue un pionero en demostrar que efectivamente la probabilidad de error podía hacerse tan pequeña como se quisiera con una tasa de transmisión estrictamente positiva, lo que no se creía posible en esa época. Además sus análisis mostraron que las tasas de transmisión empleadas en su época estaban muy por debajo de las que podrían emplearse [3].

Uno de los modelos que es usado frecuentemente para canales de comunicaciones prácticos es el canal de ruido blanco Gaussiano aditivo (Fig. 2.2) (*additive white Gaussian noise*, AWGN), cuya relación entrada-salida es:

$$Y_i = gX_i + Z_i, \quad i = 1, 2, \ldots \tag{2.8}$$

y donde $\{Z_i\}_i$ es una secuencia de variables aleatorias independientes idénticamente distribuidas normales complejas circularmente simétricas [6] de varianza σ^2, independientes de

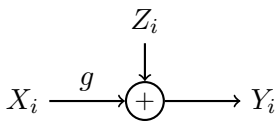

Figura 2.2: Representación del canal de ruido blanco aditivo Gaussiano (ecuación 2.8).

$\{X_i\}_i$, y g es una ganancia fija, compleja y conocida al menos al receptor. Este canal es usado habitualmente para representar un canal ideal con ruido térmico en los receptores de comunicaciones [21].

Los alfabetos de entrada y salida de éste canal son continuos, es decir, tanto la salida como la entrada podrían ser cualquier número complejo. Si se emplea un código (M, n) arbitrario como se indicó antes para transmitir, podrían elegirse secuencias de símbolos lo suficientemente alejados entre sí como para que la probabilidad de error sea tan chica como se quiera para cualquier tasa, siendo la capacidad infinita. Por lo tanto, debe incorporarse alguna restricción adicional a los códigos empleados. Dada una secuencia de entrada $(x_1, \ldots, x_n)$, habitualmente se considera una restricción de energía:

$$\frac{1}{n}\sum_{i=1}^{n}|x_i|^2 \leq P, \tag{2.9}$$

donde $P > 0$ es el valor de restricción. Shannon demostró que la capacidad de este canal es:

$$C = \sup_{X:\mathbb{E}[|X|^2]<P} I(X;Y) = \log_2(1+\mathrm{SNR}), \tag{2.10}$$

donde $\mathrm{SNR} = |g|^2P/\sigma^2$ es la relación señal a ruido en el receptor. $I(X,Y)$ es la información mutua entre las variables X e Y, definida como:

$$I(X;Y) \triangleq \mathbb{E}\left[\log_2\left(\frac{f_{X,Y}(X,Y)}{f_X(X)f_Y(Y)}\right)\right], \tag{2.11}$$

donde $f_{X,Y}(x,y)$, $f_X(x)$, $f_Y(y)$ son las funciones de densidad de las variables aleatorias indicadas. El supremo en (2.10) es sobre todas las densidades de probabilidad que satisfacen la restricción sobre la varianza. Dicho supremo se alcanza cuando X es una variable compleja normal circularmente simétrica de media nula y varianza P. La prueba del teorema de hecho muestra que la capacidad puede alcanzarse si las palabras de códigos se generan como una secuencia de variables independientes Gaussianas con varianza P; dichos códigos no son útiles desde un punto de vista práctico por la complejidad requerida en la decodificación, pero son útiles como herramientas de análisis teórico de sistemas de comunicaciones, como veremos más adelante.

Volviendo sobre el canal (2.3) de la sección anterior, se observa que para cada realización del coeficiente de *fading* g, el canal (2.3) corresponde con el canal Gaussiano (2.8). Esto implica que la capacidad será una función de la realización del *fading*:

$$C(g) = \log\left(1 + \frac{|g|^2 P}{\sigma^2}\right). \tag{2.12}$$

Dado que el coeficiente de *fading* es aleatorio y fijado antes de comenzar la transmisión, la capacidad del canal también será una variable aleatoria: antes de comenzar se selecciona el *fading* y éste define la capacidad del canal durante esa transmisión. Si se utiliza la noción de capacidad tradicional sobre el canal, debe hallarse una tasa que sea transmisible en forma segura sobre todas las realizaciones del *fading*. Dado que en general el *fading* puede ser arbitrariamente pequeño (dependiendo de su distribución), para cualquier tasa $R > 0$ seleccionada, hay una probabilidad positiva de que $C(g) < R$. Por lo tanto, de acuerdo a esta noción, la capacidad del canal *slow fading* es cero.

Esto motiva una definición diferente para la capacidad de este tipo de canales. Si se asume que el canal hacia el destino es conocido sólo en el destino y no en el transmisor, éste no puede adaptar la tasa a la realización del canal. Por lo tanto, para cada tasa posible R seleccionada, si ocurre el evento $\{C(g) < R\}$ la tasa no será alcanzable y se dirá que hay un evento de falla o *outage*. La probabilidad de *outage*, puede hallarse como:

$$P_{\text{out}}(g) \triangleq \mathbb{P}(C(g) < R) = \mathbb{P}(|g|^2 < \sigma^2(2^R - 1)/P) = F_{|g|^2}\left(\frac{\sigma^2}{P}(2^R - 1)\right), \tag{2.13}$$

que es la probabilidad de que la tasa seleccionada R no pueda ser transmitida por el canal. Esta probabilidad lleva a la definición de capacidad de outage del canal, como la máxima tasa que puede transmitirse si se permite una probabilidad de outage máxima de ϵ:

$$C_{\text{out}} = \max_{g: P_{\text{out}}(g) < \epsilon} C(g). \tag{2.14}$$

La noción de probabilidad de outage, puede ser generalizada a canales slow *fading* más generales, como veremos más adelante. Además, la probabilidad de outage, es una cota superior a la probabilidad de error del canal, y por lo tanto, una métrica de importancia.

2.4 Modelos de redes inalámbricas mediante geometría estocástica

Los trabajos posteriores a los de Shannon utilizaron sus ideas para analizar y entender escenarios más complejos, incluyendo canales con más de un receptor y/o transmisor, como el canal de acceso múltiple (*multiple-access channel*), el canal *broadcast* o el de interferencia [4].

Para muchos de estos canales la capacidad tradicional es conocida, mientras que para otros solamente son conocidas cotas inferiores y superiores a la capacidad. Aunque el análisis de redes de comunicación con muchos nodos es posible desde la perspectiva de la teoría de la información, las expresiones obtenidas son, a menudo, muy complejas o difíciles de interpretar y analizar más allá de situaciones con pocos nodos. Las redes inalámbricas consistirán típicamente de muchos usuarios, de modo que es necesario desarrollar modelos adecuados a las mismas, que permitan su análisis y la obtención de conclusiones relevantes. Según se discutió anteriormente, el medio inalámbrico es particularmente adverso a las comunicaciones, comparado a la mayoría de los medios cableados: las señales transmitidas sufren de una atenuación con la distancia, así como también de una atenuación causada por la geometría, a menudo impredecible, del ambiente donde se realizan las transmisiones. Además, típicamente habrá numerosas transmisiones simultáneas, y, dada la naturaleza omnidireccional del medio inalámbrico y las limitaciones en términos de espectro radioeléctrico, la interferencia entre las transmisiones será inevitable. En muchas casos, dicha interferencia será mayor y un efecto más nocivo para las comunicaciones que la presencia de ruido.

Consideremos una gran red con l usuarios o nodos transmisores y m nodos receptores que comparten el mismo espacio y el mismo espectro, y cuyos canales son de desvanecimiento lento como (2.3). Dicha red, en cada instante tiempo, podría representarse por medio de un sistema de ecuaciones:

$$y_k = \sum_{j=1}^{l} a_{j,k} x_j + z_k, \qquad k = 1, \ldots, m, \tag{2.15}$$

en donde y_k es el símbolo recibido por el k-ésimo receptor en el instante k, $a_{j,k}$ es la ganancia compleja de canal y x_j es el símbolo transmitido por el usuario j y z_k es el ruido Gaussiano en el receptor. Algunos de los transmisores podrían tener uno o más mensajes para transmitir, otros podrían cooperar actuando como relevos, y algunos transmisores podrían interferir con los demás. El estudio de redes inalámbricas requerirá normalmente considerar un gran número de nodos (m y n grandes), cuyas posiciones son a menudo impredecibles y afectan la calidad de sus canales y la interferencia que experimentan. Analizar una configuración específica de la red no permitirá sacar conclusiones acerca del comportamiento promedio de los usuarios, y a menudo no será posible en forma cerrada, especialmente porque en una configuración dada los canales serán todos diferentes. Además, el comportamiento de la interferencia es diferente al del ruido, en el sentido de que el ruido no se ve afectado por las acciones de los usuarios, mientras que, en el caso de la interferencia, las acciones de los usuarios que deberían favorecerlos individualmente, pueden ser contraproducentes si son realizadas por todos los usuarios de la red al mismo tiempo. Esto no ocurre cuando solamente hay ruido, aun correlacionado, en los receptores.

La teoría de los procesos puntuales es una rama de la geometría estocástica [2] dedicada al estudio de la distribuciones aleatorias de puntos en cierto espacio. Dicha teoría brinda un marco natural para el modelado y análisis de grandes redes inalámbricas donde la interferencia es una de las limitaciones principales para el desempeño [7, 8, 9]. Mediante esta teoría, muchos de los aspectos principales de una red inalámbrica pueden ser modelados, y, promediando sobre todas las configuraciones posibles de nodos, pueden estudiarse los efectos macroscópicos de, entre otras cosas, la interferencia y su correlación espacio-temporal. Además, el modelado de redes mediante procesos estacionarios implica que el desempeño promedio de todos los nodos será el mismo, de modo que, estudiando un nodo de la red, podrá caracterizarse el comportamiento global de la misma, obteniéndose expresiones cerradas y compactas que representan a toda la red.

Un proceso puntual es estructuralmente análogo a una variable aleatoria, salvo que el resultado de una realización del proceso es un conjunto numerable de puntos en un espacio, en lugar de un número real. En el contexto de las redes inalámbricas, los puntos en general estarán en el plano, es decir, en $\mathbb{R}^2$. Normalmente se denota a un proceso puntual como $\Phi = \{u_i\}_{i\in\mathbb{N}}$, donde $u_i \in \mathbb{R}^2$ la posición del punto i-ésimo del proceso. El proceso puntual más utilizado y estudiado es el proceso de Poisson homogéneo o estacionario; dicho proceso es útil para representar situaciones en las que los nodos están distribuidos en forma uniforme en el espacio.

Definition 2.4.1 (Proceso de Poisson homogéneo)**.** *Se dice que Φ es un proceso de Poisson homogéneo de intensidad $\lambda > 0$ en $\mathbb{R}^2$ si satisface las dos condiciones siguientes:*

- *Dado un conjunto acotado A, la cantidad de puntos que el proceso posee en A, denotado como $\Phi(A)$ es una variable aleatoria Poisson de media $\lambda|A|$, donde $|\cdot|$ denota el área usual.*
- *Si $A_1, \ldots, A_n$ son conjuntos disjuntos y acotados, entonces $\Phi(A_1), \ldots, \Phi(A_n)$ son variables aleatorias independientes.*

De esta definición puede demostrarse que, dado un recinto acotado A, condicionando respecto de $\Phi(A)$, los puntos del proceso están distribuidos en forma uniforme e independiente en A. Esto permite la simulación del proceso de Poisson en un recinto arbitrario del plano, primero generando una variable Poisson de media $\lambda|A|$ y luego tirando esa misma cantidad de puntos uniformes en el recinto. La intensidad o densidad λ tiene unidades de puntos por unidad de área, y representa la densidad espacial promedio de nodos de la red.

Ilustraremos la forma de trabajar con un proceso puntual mediante un ejemplo que contiene las dificultades y aspectos principales del problema [10]. Suponga que en una red de

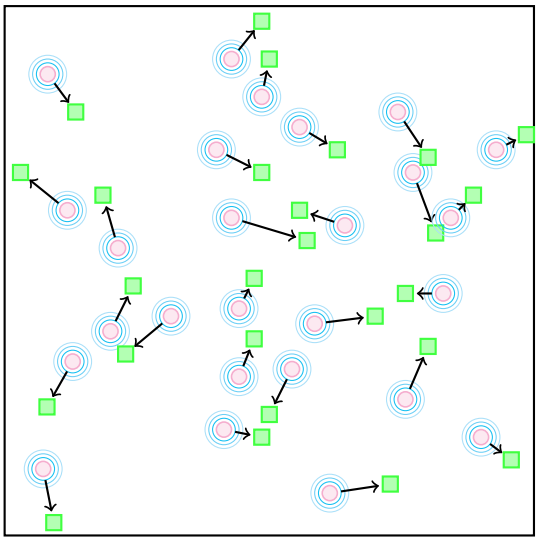

Figura 2.3: Representación de una red espacial donde los transmisores (círculos magenta) se comunican con un destino (cuadrados verdes) por medio de una comunicación punto a punto.

comunicaciones (2.15) los transmisores están distribuidos en el plano de acuerdo a un proceso puntual estacionario $\Phi = \{u_i\}$, y que cada uno de ellos tiene un destino diferente que no pertenece al proceso, con el que desea comunicarse por medio de una comunicación punto a punto (ver Fig. 2.3). Suponga que el canal es *slow fading* como (2.3), con el *fading* de media unitaria, y que el destino conoce el canal que lo separa de la fuente, pero no así la fuente. De momento no consideraremos el ruido. Nos enfocamos en un transmisor, que asumimos ubicado en el origen o, cuyo destino está ubicado en la posición d. La señal recibida en el receptor en el instante k es:

$$y_{d,k} = ||d||^{-\frac{\alpha}{2}} h_d x_{o,k} + \sum_{u \in \Phi \setminus \{o\}} ||u - d||^{-\frac{\alpha}{2}} h_{u,d} x_{u,k}, \tag{2.16}$$

donde $\setminus$ denota que hemos removido el punto del origen del proceso y por lo tanto de la sumatoria. El primer término es la transmisión del usuario en el origen, y el segundo término son las transmisiones de los otros usuarios del proceso sin contar el del origen. Este canal es una canal de interferencia donde varios usuarios intentan transmitir a la vez, generando interferencia entre ellos. En general, se asume que el receptor no intenta procesar la interferencia, sino que ésta es tomada como un ruido. De esta forma el canal se analiza como si fuera un canal punto a punto en lugar de un canal de interferencia. Observar que la interferencia depende de la estrategia de codificación que utilicen los usuarios, de modo que el análisis es diferente y mucho más complicado que el del canal AWGN que se mencionó en la sección anterior, donde el ruido es estático. Si bien no puede determinarse la mejor tasa

que puede alcanzarse cuando se considera la interferencia como ruido, sí puede determinarse una tasa que es alcanzable por el usuario, si se utiliza la misma estrategia de codificación que se utiliza en el canal AWGN. Si las palabras de código de los usuarios son generadas como realizaciones de variables Gaussianas independientes circularmente simétricas de varianza P, para cada realización del proceso puntual las señales temporal de interferencia tendrá también distribución Gaussiana, y su varianza será:

$$I(d) = \sum_{u \in \Phi \setminus \{o\}} P||u - d||^{-\alpha}|h_{u,d}|^2. \tag{2.17}$$

Dicha varianza es la potencia de la interferencia en el punto d, y será una variable aleatoria cuando se considera la aleatoriedad del *fading* y del proceso puntual. De este modo, la relación señal-interferencia (*signal-to-interference ratio*, SIR) puede definirse como el cociente entre la potencia el transmisor y la de la interferencia:

$$\mathrm{SIR}(o, d) = \frac{||d||^{-\alpha}|h_d|^2}{\sum_{u \in \Phi \setminus \{o\}} ||u - d||^{-\alpha}|h_{u,d}|^2}. \tag{2.18}$$

El transmisor podrá entonces alcanzar cualquier tasa R que satisfaga:

$$R < \log_2 \left(1 + \mathrm{SIR}(o, d)\right). \tag{2.19}$$

El problema es similar al analizado en el caso del canal AWGN en el sentido de que la tasa del lado derecho de (2.19) es una variable aleatoria, pero diferente en el sentido de que, para cada realización del proceso y del *fading*, dicha tasa no es necesariamente la capacidad del canal. El enfoque de *outage* seguido entonces para el canal AWGN es aplicable aquí. Como el transmisor no tiene acceso a la SIR en el receptor, no puede elegir R para adaptarse al canal; por lo tanto, diremos que ocurrirá un evento de *outage* siempre que (2.19) no se cumpla[1].

Para evaluar la probabilidad de outage es importante observar que debe promediarse respecto del *fading* de todos los usuarios, así como también respecto del proceso puntual, es decir, de las posiciones de todos los nodos. Es importante observar que el transmisor en el origen, que forma parte del proceso, está fijo y por lo tanto, se no se está promediando respecto de la distribución del proceso a secas, sino respecto de todas realizaciones del proceso que posean un nodo en el origen. Esto es similar a promediar respecto de la distribución del proceso condicional a que hay un nodo en el origen. Dado que la probabilidad de que haya un nodo en cualquier punto del plano es nula, la definición de esta distribución no puede hacerse en el

[1]En realidad, el evento de *outage* más pequeño, implicaría buscar la mejor estrategia de codificación para cada realización del proceso puntual y del *fading*, dando la mejor tasa alcanzable en cada caso. Luego, ocurriría un outage cuando la realización del *fading* y del proceso puntual impida que esas tasas sean alcanzables. Trabajando con (2.19), para cada realización del proceso y el *fading*, la tasa alcanzable es siempre menor que la óptima, dando como resultado una cota superior a la probabilidad de *outage* óptima del sistema.

sentido tradicional de la probabilidad condicional, como un cociente de probabilidades. Esto conduce a definir las llamadas distribuciones de Palm de un proceso puntual, que justamente cumplen el rol las distribuciones condicionales. Para hallar la probabilidad de outage P_{out} debemos calcular entonces:

$$P_{\text{out}} = \mathbb{P}^0 \left(R < \log_2 \left(1 + \text{SIR}(o, d) \right) \right), \tag{2.20}$$

$$= 1 - \mathbb{E}^0 \left[F_{|h_d|^2} \left(\frac{(2^R - 1) I(d)}{||d||^{-\alpha}} \right) \right], \tag{2.21}$$

donde $F_{|h_d|^2}$ es la distribución del *fading* y $\mathbb{P}^0$, conocida como la medida de probabilidad de Palm asociada al proceso, promedia respecto de todas las realizaciones del proceso que tienen un punto en el origen. La probabilidad interna es respecto del coeficiente de *fading*. Si el *fading* es Rayleigh de media unitaria, $|h_d|^2$ tiene distribución exponencial, y entonces (2.21) queda:

$$P_{\text{out}} = 1 - \mathbb{E}^0 \left[\exp \left\{ \frac{(2^R - 1) I(d)}{||d||^{-\alpha}} \right\} \right] \triangleq \mathcal{L}^0_{I(d)} \left(\frac{(2^R - 1)}{||d||^{-\alpha}} \right), \tag{2.22}$$

donde $\mathcal{L}^0_{I(d)}$ es la transformada de Laplace de la variable aleatoria interferencia, respecto de la probabilidad de Palm del proceso. En general, las distribuciones de Palm de los procesos puntuales son difíciles de hallar o no tienen forma cerrada. El proceso de Poisson se caracteriza porque su distribución de Palm coincide con la distribución normal de un proceso de Poisson (sin condicionar) a la que se le agrega un punto en el origen (ver Teorema 4.6). Además, para el caso del proceso de Poisson, la transformada de Laplace admite una forma cerrada (ver Lema 4.3), obteniéndose:

$$P_{\text{out}} = 1 - \exp \left\{ -\lambda \tilde{C} (2^R - 1)^{2/\alpha} ||d||^2 \right\}, \tag{2.23}$$

donde $\tilde{C}$ es una constante que depende sólo de α. De hecho, la estacionariedad del proceso puntual implica que dicha probabilidad de outage no depende de dónde está ubicado el transmisor, y por lo tanto representa a cualquier transmisor de la red, y es una métrica global de la red.

El ejemplo que acabamos de desarrollar ilustra varias de las ideas principales utilizadas habitualmente en el modelado de grandes redes por medio de geometría estocástica. Esencialmente, la determinación de las tasas alcanzables, sus eventos de outage asociados, las distribuciones de Palm y las ventajas de los procesos de Poisson. En lo que sigue presentaremos las contribuciones principales del presente trabajo, apoyándonos en lo desarrollado hasta ahora. Una descripción más detallada de las definiciones y resultados esenciales acerca de procesos puntuales puede encontrarse en el Capítulo 4.

2.5 Contribuciones: modelado de redes cooperativas

El concepto de comunicación cooperativa [11, 12], en el cual uno o más usuarios comparten sus antenas para generar un arreglo distribuido que les permita mejorar la tasa y la confiabilidad de sus transmisiones, surgió hace varios años con el objetivo de satisfacer las demanda creciente de tráfico y calidad de servicio de las redes inalámbricas. En general, el estudio de comunicaciones cooperativas se realiza estudiando un solo canal aislado, donde una de las limitaciones principales consideradas es la presencia de ruido aditivo descorrelacionado en los receptores. En este contexto, la cooperación parece brindar ganancias importantes en calidad de servicio. Sin embargo, los canales inalámbricos muchas veces no pueden ser considerados en forma aislada, y por lo tanto, es necesario considerar los efectos que pueden tener los usuarios vecinos. Además, si los vecinos cooperan para transmitir o retransmitir los mensajes de otros usuarios sin coordinación o control alguno, los niveles de interferencia pueden crecer hasta que los efectos benéficos de la cooperación dejen de ser sustanciales. El objetivo global del presente trabajo es el estudio de algunos esquemas cooperativos en el contexto de grandes redes inalámbricas desde la perspectiva de la geometría estocástica, es decir, considerando los efectos adversos de la interacción entre los usuarios mediante la interferencia. En las secciones que siguen resumiremos el contenido de los capítulos de esta tesis, indicando en cada caso las contribuciones principales.

2.5.1 Acerca de la utilización de relevos en redes inalámbricas

Uno de los primeros ejemplos de un canal cooperativo es quizá el canal de relevo (*relay channel*) [13], en el que un transmisor se comunica con su destino con la ayuda de un nodo adicional, conocido como el relevo (*relay*). En la Figura 2.4 puede verse una representación del canal de *relay* Gaussiano, que es similar al canal punto a punto de la Figura 2.2, con la excepción de la presencia del *relay*, que forma un canal paralelo para cooperar con la fuente. Las ecuaciones del canal de *relay* Gaussiano son:

$$Y_{r,i} = g_{sr} X_{s,i} + Z_{r,i} \tag{2.24}$$

$$Y_{d,i} = g_{sd} X_{s,i} + g_{rd} X_{r,i} + Z_{d,i}, \qquad i = 1, 2, \ldots, \tag{2.25}$$

donde los ruidos $\{Z_r\}$ y $\{Z_d\}$ son secuencias de ruido blanco complejo Gaussiano, con distintas varianzas en general e independientes entre sí y de las transmisiones de los usuarios. $X_{s,i}$ y $X_{r,i}$ son los símbolos de la fuente y el *relay* en el instante i. Las esquemas de codificación principales para el canal de *relay* fueron introducidas en [14]: decodificar-y-reenviar (*decode-and-forward*, DF) y comprimir-y-reenviar (*compress-and-forward*, CF). En la primera, el *relay* decodifica el mensaje enviado por la fuente, lo recodifica y reenvía al destino, que decodifica

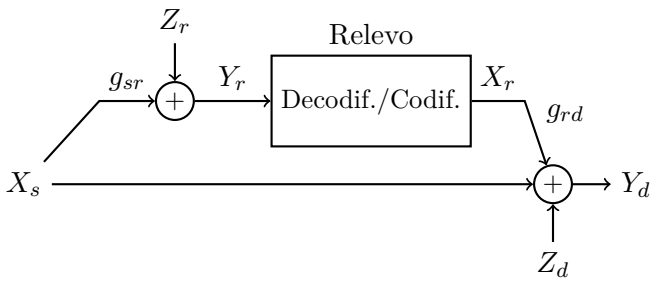

Figura 2.4: Modelo matemático del canal de relevo.

el mensaje utilizando tanto la transmisión de la fuente como del *relay*. En CF, el *relay* no decodifica el mensaje, sino que busca en una librería un mensaje que mejor se vincule con la secuencia recibida de la fuente. Luego envía el índice de dicho mensaje al destino, que conoce también la librería, y utiliza el índice como un ayuda para decodificar el mensaje junto con la transmisión de la fuente. En este sentido, el índice enviado actúa como una compresión del mensaje de la fuente, y el tamaño de la librería indica el grado de compresión del mensaje. Considerando un solo canal de *relay* aislado y en presencia de ruidos descorrelacionados en los receptores se han observado ganancias importantes con respecto a la comunicación sin *relay*, es decir, respecto de un canal punto a punto [15, 16]. En los capítulos 5 y 6 se extiende dicho estudio de estos esquemas en el contexto de una gran red inalámbrica.

En el Capítulo 5 se estudia el desempeño, en términos de la probabilidad de *outage*, de un canal de *relay* full-duplex cuando es sometido a la interferencia de una gran red inalámbrica en la que los nodos están distribuidos como un proceso de Poisson homogéneo en el plano $\Phi = \{x_i\}_i$ de intensidad λ. Este análisis se diferencia de la situación habitual en la que los receptores (*relay* y destino) hay ruidos blancos y descorrelacionados. Las señales de interferencia en los receptores están correlacionadas porque están generadas por el mismo conjunto de nodos, y, además, dicha correlación es aleatoria por la distribución espacial aleatoria de los nodos que la generan. El *relay* opera en modo *full-duplex* cuando puede recibir el mensaje de la fuente y transmitir al mismo tiempo. Se consideran *half-duplex* los protocolos en los que el *relay* escucha una fracción de tiempo y luego transmite la fracción restante.

Suponiendo que la fuente está en el origen, el relevo en r y el destino en d, y siguiendo la ecuación (2.16) del canal punto a punto, las señales en el destino y el relevo pueden escribirse (omitiendo el índice del tiempo) como:

$$Y_r = h_{sr}\sqrt{l_{sr}}X_s + \underbrace{\sum_{x\in\Phi} h_{xr}l(x,r)^{\frac{1}{2}}X_x}_{\tilde{Z}_r}$$

$$Y_d = h_{sd}\sqrt{l_{sd}}X_s + h_{rd}\sqrt{l_{rd}}X_r + \underbrace{\sum_{x\in\Phi} h_{xd} l(x,d)^{\frac{1}{2}} X_x}_{\tilde{Z}_d},$$

donde $\{X_{x_i}\}_i$ son los símbolos transmitidos por los otros transmisores de la red. El modelo de atenuacion es (2.3) y l_{sd}, l_{rd}, l_{sr} son las atenuaciones por las distancias entre la fuente y el destino, *relay* y destino, y fuente y *relay*, repectivamente. Si los códigos empleados por los usuarios son Gaussianos, igual que en el caso punto a punto, y de varianza unitaria y la interferencia es tratada como ruido, para cada realización del proceso puntual, las señales de interferencia temporales $\tilde{Z}_r$ and $\tilde{Z}_d$, son variables Gaussianas complejas circularmente simétricas de media nula, y varianzas:

$$I_r \triangleq \mathbb{E}\left[|\tilde{Z}_r|^2|\tilde{\Phi}\right] = \sum_{x\in\Phi} |h_{xr}|^2 l(x,r), \tag{2.26}$$

$$I_d \triangleq \mathbb{E}\left[|\tilde{Z}_d|^2|\tilde{\Phi}\right] = \sum_{x\in\Phi} |h_{xd}|^2 l(x,d). \tag{2.27}$$

Además, las señales temporales están correlacionadas (porque son generadas por los mismos transmisores), y su coeficiente de correlación vale:

$$\rho_N = \frac{\mathbb{E}[\tilde{Z}_r \tilde{Z}_d^*|\tilde{\Phi}]}{\sqrt{I_r I_d}}. \tag{2.28}$$

Utilizando los códigos Gaussianos pueden hallarse las tasas que pueden alcanzarse por DF y CF, y determinar los eventos de *outage* de cada uno de los protocolos. Para el caso de DF, es posible que si el *relay* no decodifica el mensaje de la fuente, el destino igual pueda decodificar la transmisión de la fuente, como si el *relay* no existiera. Además, la tasas que son alcanzables para DF no dependen explícitamente de la correlación (2.28), por el requisito de que el *relay* decodifique el mensaje de la fuente en forma completa, lo que separa ambas transmisiones. Por este motivo, es posible hallar una expresión cerrada de la probabilidad de *outage* (ver Sección 5.3.1 para los detalles):

Teorema 2.1 (OP de DF)**.** *La probabilidad de* outage *de DF se escribe como:*

$$P_{out,DF}(R,\rho) = 1 - \mathbb{P}\left(\mathcal{A}_{DF}^c(R,\rho) \cap \mathcal{B}_{DF}^c(R,\rho)\right) - \mathbb{P}\left(\mathcal{A}_{DF}(R,\rho) \cap \mathcal{A}_{DT}^c\right), \tag{2.29}$$

donde $\mathcal{A}_{DF}$ es el evento de que el relay *no decodifique la transmisión de la fuente, $\mathcal{B}_{DF}$ que el destino no decodifique la transmisión conjunta del* relay *y la fuente, y $\mathcal{A}_{DT}$, que el destino no decodifique la transmisión del source (utilizada en caso de que el* relay *falle como se mencionó anteriormente). R es la tasa a la que intenta transmitir la fuente, y ρ es el coeficiente de correlación entre los símbolos de la fuente y del* relay*. Los términos de (2.29) cuando $||r-d|| \neq ||d||$ o $\rho \neq 0$ se escriben como:*

$$\mathbb{P}\left(\mathcal{A}_{DF}^c(R,\rho) \cap \mathcal{B}_{DF}^c(R,\rho)\right) = \frac{\mu_2 \mathcal{L}_{I_d,I_r}\left(\frac{T(R)}{\mu_2}, \frac{T(R)}{\mu_3}\right) - \mu_1 \mathcal{L}_{I_d,I_r}\left(\frac{T(R)}{\mu_1}, \frac{T(R)}{\mu_3}\right)}{\mu_2 - \mu_1}, \tag{2.30}$$

donde:

$$\mu_1 = \frac{1}{2}\left[l_{sd} + l_{rd} - \left((l_{sd} - l_{rd})^2 + 4l_{sd}l_{rd}|\rho|^2\right)^{\frac{1}{2}}\right], \tag{2.31}$$

$$\mu_2 = \frac{1}{2}\left[(l_{sd} + l_{rd}) + \left((l_{sd} - l_{rd})^2 + 4l_{sd}l_{rd}|\rho|^2\right)^{\frac{1}{2}}\right], \tag{2.32}$$

$$\mu_3 = l_{sr}\left(1 - |\rho|^2\right), \tag{2.33}$$

y $\mathcal{L}_{I_d,I_r}(\omega_1, \omega_2)$*, es la transformada de Laplace conjunta de las interferencias, dada por (5.5),* $y\ T(R) = 2^R - 1$*. Además, cuando* $||r - d|| = D$ *y* $\rho = 0$*, se tiene que* $\mu_1 = \mu_2$ *y:*

$$\mathbb{P}\left(\mathcal{A}_{DF}^c(R,\rho) \cap \mathcal{B}_{DF}^c(R,\rho)\right) = \mathcal{L}_{I_d,I_r}\left(\frac{T(R)}{\mu_1}, \frac{T(R)}{\mu_3}\right) - \frac{T}{\mu_1}\frac{d\mathcal{L}_{I_d,I_r}(\omega_1, T(R)/\mu_3)}{d\omega_1}\bigg|_{\omega_1 = T(R)/\mu_1}. \tag{2.34}$$

Finalmente:

$$\mathbb{P}\left(\mathcal{A}_{DF} \cap \mathcal{A}_{DT}^c\right) = \mathcal{L}_{I_d}\left(\frac{T(R)}{l_{sd}}\right) - \mathcal{L}_{I_d,I_r}\left(\frac{T(R)}{l_{sd}}, \frac{T(R)}{\mu_3}\right). \tag{2.35}$$

Se demuestra, además, (Lema 5.3) que, en el régimen de operación típico de las redes inalámbricas, el coeficiente de correlación ρ entre los símbolos que transmiten la fuente y el relevo es $\rho = 0$, lo que simplifca las expresiones del teorema anterior.

El análisis de CF es más complejo que el de DF, pues en CF el *relay* comprime la información sin decodificar el mensaje, y esto implica que la correlación (2.28) aparece explícitamente en las tasas alcanzables. Por lo tanto, el evento de *outage* de CF, denotado como $\mathcal{O}_{CF}(R, n_c, \rho_N)$, dependerá de la tasa R que intenta la fuente, de la correlación entre las interferencias ρ_N dada por (2.28), y también de n_c que es un parámetro positivo de diseño que controla cuánto *comprime* el *relay* el mensaje de la fuente. Dado que el comportamiento de dicha correlación depende de la distribución del proceso, no puede hallarse una expresión cerrada para la probabilidad de *outage*. Para poder hallar una cota de la probabilidad de *outage*, en el Lema 5.6 se demuestra que la tasa alcanzable en CF para una correlación arbitraria de la interferencia ρ_N es al menos un bit peor que con interferencia descorrelacionada. Esto implica que puede acotarse la probabilidad de *outage* de CF, aumentando la tasa y asumiendo que la interferencia está descorrelacionada (detalles en la Sección 5.3.2):

Teorema 2.2 (OP de CF)**.** *Cuando la fuente intenta comunicarse con el destino a una tasa* R*, la probabilidad de* outage *de CF puede acotarse como:*

$$\begin{aligned} P_{out,CF}(R, n_c) &= \mathbb{P}(\mathcal{O}_{CF}(R, n_c, \rho_N)) \\ &\leq \mathbb{P}(\mathcal{O}_{CF}(R+1, n_c, 0)) \\ &= \mathbb{P}(\mathcal{A}_{CF}(R+1, 0) \cup \mathcal{B}_{CF}(n_c, 0)), \end{aligned} \tag{2.36}$$

donde $\mathcal{A}_{CF}$ es el evento de que falle la transmisión entre el relay *y el destino, es decir, de haber elegido un valor de n_c que sea demasiado bajo como para permitir la transmisión de la secuencia comprimida. $\mathcal{B}_{CF}$ es el evento de que el destino no pueda decodificar el mensaje con ambas transmisiones. Además, la probabilidad de los eventos con $\rho_N = 0$ puede acotarse como:*

$$\mathbb{P}(\mathcal{A}_{CF}(R,0)) \leq 1 - e^{-\frac{T(R)n_c}{l_{sr}}}\mathcal{L}_{I_r}\left(\frac{T(R)}{l_{sr}}\right) - \left[\sum_{\tilde{n}=0}^{\tilde{N}-1} e^{\frac{\tilde{n}n_cT(R)}{\tilde{N}l_{sr}}}\mathcal{L}_{I_d,I_r}\left(\frac{(\tilde{N}-\tilde{n})T(R)}{\tilde{N}l_{sd}}, \frac{\tilde{n}T(R)}{\tilde{N}l_{sr}}\right)\right.$$
$$\left. - e^{\frac{(\tilde{n}+1)n_cT(R)}{\tilde{N}l_{sr}}}\mathcal{L}_{I_d,I_r}\left(\frac{(\tilde{N}-\tilde{n})T(R)}{\tilde{N}l_{sd}}, \frac{(\tilde{n}+1)T(R)}{\tilde{N}l_{sr}}\right)\right], \quad (2.37)$$

y:

$$\mathbb{P}(\bar{\mathcal{A}}_{CF}(R,0) \cap \mathcal{B}_{CF}(n_c,0)) \leq$$
$$1 - \mathbb{E}\left[\mathcal{L}_{I_d,I_r}\left(\frac{(1+T(R))l_{sr}|h_{sr}|^2}{T(R)n_cl_{rd}}, \frac{(1+T(R))l_{sd}|h_{sd}|^2}{T(R)n_cl_{rd}}\right)\right]. \quad (2.38)$$

Esta última esperanza es respecto de h_{sr} y h_{sd}, con $T(R) = 2^R - 1$.

Finalmente, luego de caracterizar los dos protocolos principales, se comparan sus probabilidades de *outage* y las cotas para la capacidad de *outage* que proveen, cuando se los compara con la transmisión punto a punto, y contra la versión half-duplex de DF, en el que el *relay* escucha la transmisión de la fuente una fracción de tiempo y luego retransmite la fracción restante. En este último caso se provee una cota inferior de la probabilidad de *outage* (Teorema 5.5), que pueda ser comparada con las cotas superiores de los otros protocolos. En la Sección 5.4 se presentan los resultados numéricos que comparan los protocolos estudiados. En particular, mencionaremos la Figura 5.7, donde se grafica la máxima tasa que puede alcanzarse con cada uno de los protocolos cuando se requiere que la probabilidad de *outage* sea inferior a 0,05, cuando el *relay* está sobre la línea que separa la fuente del destino. Puede observarse que la versión *full-duplex* de DF funciona mejor que la versión *half-duplex*, especialmente cuando el *relay* está a distancia similar de la fuente y del destino. No debe olvidarse que para el caso *half-duplex*, se provee una cota inferior de la probabilidad de *outage*, lo que da una cota superior de la máxima tasa alcanzable. Esto implica que en realidad podría ser aun mayor la diferencia entre ambos protocolos. Por otro lado, se ve que CF no logra ser mejor que DF en ningún caso, lo que sugiere que las cotas desarrolladas suponiendo que la correlación es nula pueden ser muy conservativas, es decir, que los efectos de la correlación son importantes. Por este motivo, en la Figura 5.8 se comparan las regiones del espacio en las que CF, DF full-duplex o transmisión directa son más convenientes. Para el caso de CF, se realiza una simulación Monte Carlo de la probabilidad de *outage*, en la que se cuenta la

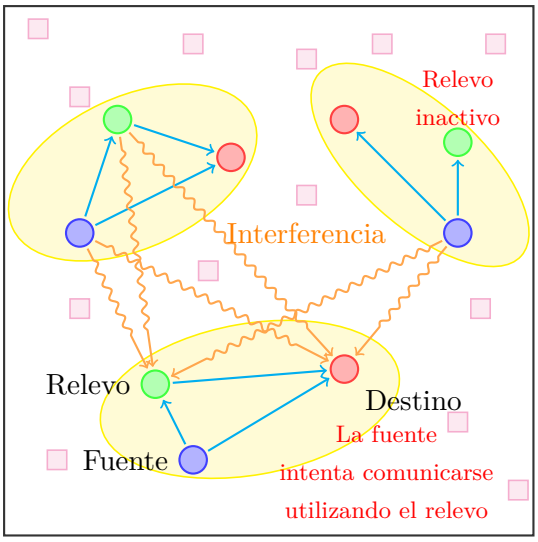

Figura 2.5: La red está formada por grupos de usuarios (en amarillo) que utilizan DF o DT. Los relevos (en verde) son elegidos del conjunto de los nodos inactivos de la red (cuadrados magenta) para asistir a las fuentes (azules) en su transmisión hacia sus destinos (en rojo).

correlación. Se observa que alrededor del destino hay una región en la que CF es mejor que los demás protocolos, pero que DF domina en la mayor parte de la región.

Luego de estudiar algunos aspectos del desempeño de un canal de *relay* cuando la interferencia es generada por un proceso puntual, en el Capítulo 6 se procede a estudiar una situación en la que también algunos usuarios que generan interferencia puedan utilizar un relevo. En particular, se considera una red decentralizada en la que se utiliza el protocolo DF *full-duplex*, y se considera que cada transmisor de la red puede utilizar un *relay*, que es elegido como el nodo inactivo más cercano de la fuente en la dirección del destino (ver Figura 2.5). Se introduce un esquema de activación sencillo en el cual los *relays* se activan en forma independiente, y que puede representar la disponibilidad de los relevos para cooperar con sus fuentes. Las expresiones son ahora mucho más complicadas pues habrá términos de interferencia que contengan más de un transmisor y otros términos que contengan uno sólo. De todas formas, puede hallarse la probabilidad de *outage* del protocolo (Sección 6.3):

Teorema 2.3. *La probabilidad de* outage *de la red en la que algunos usuarios poseen relevos*

$P_{out,mix}$ vale:

$$P_{out,mix}(R) = \mathbb{P}(\varepsilon_0 = 0)\left[1 - \mathcal{L}_{I_d}\left(\frac{T(R)}{l_{sd}}\right)\right] + \mathbb{P}(\varepsilon_0 = 1) \times \mathbb{E}_r\left[\frac{D^\alpha \mathcal{L}_{I_d,I_r}\left(\frac{T(R)}{l_{rd}}, \frac{T(R)}{l_{sr}}\right) - ||r-d||^\alpha \mathcal{L}_{I_d,I_r}\left(\frac{T(R)}{l_{sd}}, \frac{T(R)}{l_{sr}}\right)}{D^\alpha - ||r-d||^\alpha}\right], \quad (2.39)$$

donde, al igual que en casos previos, $\mathcal{L}_{I_d,I_r}(\cdot,\cdot)$ es la transformada de Laplace de las interferencias, cuya expresión general viene dada por (5.5). La variable aleatoria Bernoulli ε_0 indica que la fuente bajo estudio tiene un relay *que coopere con ella o no.*

Luego se procede a estudiar cuál es la probabilidad de activación de los relevos que minimiza la probabilidad de *outage* de los usuarios. Para ello, se considera el caso en el que la distribución de los *relays* está concentrada alrededor de la fuente, y dentro de valores prácticos para la operación de la red. En estas circunstancias se demuestra (Teorema 6.3) que, bajo este esquema sencillo, la mejor elección es activar o no todos los relevos al mismo tiempo, y esta elección depende de los parámetros fundamentales de la red, como el exponente de atenuación del *path loss*, la distancia al destino, la densidad de fuentes, la tasa, etc. Esto no es cierto, sin embargo, cuando la probabilidad de *outage* se encuentra en valores grandes, es decir fuera del régimen de operación habitual de una red. Con estos resultados, se concluye el estudio de la utilización de relevos en grandes redes inalámbricas y los efectos combinados de la cooperación y la interferencia.

2.5.2 Acerca de las comunicaciones entre usuarios móviles en redes inalámbricas

El tráfico en las redes celulares he experimentado un incremento notable en los últimos años, motivado principalmente por la transición de los sistemas de comunicación puramente de voz (2G y anteriores) a sistemas con predominancia de tráfico de datos (3G en adelante). Esto hace necesario contar con técnicas novedosas y diversas para satisfacer dichas demandas, entre ellas, un mejor aprovechamiento del espectro, mejores esquemas de modulación, reducir las distancias de transmisión, etc. De todas ellas, la que ha exhibido mayores ganancias es la reducción del tamaño de las celdas y de las distancias de transmisión dentro de la red [17], conduciendo a una mayor reutilización del espectro y a una mayor eficiencia espectral. Estas conclusiones introducirán cambios en la infraestructura de las redes celulares, que cambiarán de un paradigma homogéneo, centrado en estaciones base dispersas de largo alcance, a un paradigma heterogéneo, con una red densa de bases con distinto radio de cobertura, y considerando interacciones más complejas como la cooperación entre estaciones base y entre usuarios.

Una de las posibles estrategias, que serán empleadas en el contexto de redes móviles para lograr este objetivo, es el de las comunicaciones entre móviles (*device-to-device*, D2D) [18], en la que dos usuarios móviles con proximidad espacial puede conectarse en forma directa mediante una transmisión de corta distancia, evitando un salto por medio de la estación base [19]. Esta estrategia contribuirá a la reutilización de las frecuencias, a la eficiencia energética y a reducir la latencia, entre otras cosas. Las estimaciones actuales [1] de los incrementos de tráfico para los próximos años indican que una de las causas principales del incremento de tráfico tendrán que ver con la transmisión de videos bajo demanda (*on-demand*), en las que un gran número de usuarios accede a una librería pequeña de videos, de un modo asincrónico, de modo que una única transmisión no puede usarse para todos los dispositivos [20]. En este contexto, la gran cantidad de memoria de almacenamiento libre que tienen disponibles la mayorí a de los móviles de hoy en día podría utilizarse para almacenar en forma distribuida contenidos que luego pueden ser distribuidos a través de D2D, en lugar de ser descargados de la estación base. Además, si los intercambios ocurren fuera de la banda de frecuencia de telefonía móvil *out-of-band D2D*, aumenta la reutilización de frecuencia.

El objetivo del Capítulo 7 es estudiar algunos aspectos del desempeño de estrategias de intercambio de videos D2D cuando los videos se intercambian fuera de la banda de frecuencia, utilizando para ello un modelo de procesos puntuales. En particular, se asume que en un determinado instante en una gran red espacial hay usuarios distribuidos que quieren videos y otros que los tienen almacenados, luego de haberlos visto previamente. Se desea estudiar la fracción de los pedidos de videos que podrían ser satisfechas mediante D2D, es decir, mediante transmisiones locales, en lugar de pedir una transmisión por parte de la estación base. Esta fracción es un indicador de las potenciales ventajas económicas y en términos de calidad de servicio que pueden obtenerse por medio de D2D. Dicho problema no puede ser estudiado analizando solamente si las transmisiones tienen éxito, pues la posibilidad de que un archivo pedido esté o no disponible en la cercanía de un usuario debe ser tenida en cuenta. De hecho, el problema tiene tres aspectos principales:

- La estrategia de almacenamiento de videos en los usuarios, es decir, qué videos almacenan los usuarios, cuántos almacenan, etc.
- El problema de vincular los usuarios entre sí para que puedan intercambiar archivos, es decir, cómo decidir qué usuario le transmite a quién el video pedido.
- Los problemas asociados con la transmisión y la coordinación entre los usuarios, es decir, cómo hacer las transmisiones y las posibilidades de falla de las mismas.

Como primera contribución del trabajo, se introduce un modelo de procesos puntuales de la

red, brindando un marco teórico para el análisis de estrategias D2D. Dicho modelo tiene las siguientes características:

- Se considera la existencia de usuarios que piden videos (destinos) y de usuarios cooperativos, (con videos almacenados) que se distribuyen en el espacio como procesos de Poisson homogéneos e independientes. Las comunicaciones entre los usuarios se modelan utilizando el modelo *slow fading* dado por (2.3).

- Los usuarios son agrupados en el espacio, en grupos denominados *clusters*, y dentro de esos *clusters* ocurren los intercambios D2D de videos. Se introduce una familia de *protocolos admisibles* que son utilizados para coordinar la red. Cada protocolo está compuesto por una estrategia para formar los *clusters*, que es inducida por un proceso puntual de núcleo duro[2] (*hard core*) y una estrategia de coordinación dentro de un *cluster*, que indica cómo se comportan los usuarios dentro de los *clusters*.

En este contexto, se definen dos métricas de interés, que caracterizan la cantidad de pedidos que pueden ser satisfechos mediante D2D:

- Una *métrica global*, definido como el cociente entre la densidad de pedidos satisfechos mediante D2D y la cantidad total de pedidos de videos en cualquier punto de la red.

- Una *métrica local*, que indica el cociente entre el número promedio de pedidos satisfechos y el número promedio de pedidos dentro de un *cluster*.

Aunque estas métricas apuntan a estudiar la cantidad de pedidos que pueden atenderse mediante D2D, y consideran los tres aspectos mencionados del problema mencionados, no se consideran requisitos en cuanto a la calidad de los enlances que puede requerirse en el intercambio de videos. Es decir, podría satisfacerse un gran número de pedidos, pero con un tasa de transmisión demasiado baja como para satisfacer a los usuarios. Por este motivo, se introducen tres regiones de interés, donde se consideran las dos métricas definidas considerando requisitos de tasas promedio para los enlaces:

- Una región donde se maximiza sobre todos los protocolos admisibles la métrica global, ponderada por un requisito de tasa promedio por usuario.

- Una región donde se maximiza sobre todos los protocolos admisibles la métrica local, ponderada por un requisito de tasa promedio por usuario.

[2]Un proceso puntual de *hard core* es un proceso que garantiza una distancia mínima entre los puntos. Se utiliza en este caso con el objetivo de garantizar que los clusters no se intersequen entre sí, y que cada usuario pertenezca a un sólo *cluster*.

- Una región de compromiso donde se ponderan ambas métricas, es decir, se maximiza sobre todos los protocolos admisibles la fracción local y global de usuarios que pueden atenderse a la vez.

Dado que no pueden determinarse el o los protocolos que definen estas regiones, puede analizarse un protocolo en particular, para hallar una cota interior a las regiones. Para obtener cotas interiores de dichas regiones, se introduce entonces una estrategia de coordinación que pueda ser analizada; dicha estrategia puede formar un protocolo admisible con cualquier estrategia de agrupado de las mencionadas anteriormente. En la estrategia propuesta, los usuarios que tienen y piden videos son apareados y los videos se intercambian mediante transmisiones punto a punto. Dentro de los clusters la interferencia se reduce por medio de la división del canal en ranuras de tiempo, ocupada cada una por una transmisión. Al analizar la estrategia, se observa que la interferencia durante las transmisiones es no estacionaria, requiriendo un análisis adicional para considerar este hecho. Con el análisis de esta estrategia se hallan cotas interiores para las regiones definidas, sacando conclusiones acerca de la desempeño alcanzable mediante D2D.

On Large Cooperative Wireless Network Modeling through a Stochastic Geometry Approach

Chapter 3

Introduction

3.1 Preliminaries

Wireless communications are the fastest growing segment in the communications industry [3]. New applications are continuously emerging, converging to a vision in which most embedded devices on Earth will be connected to a network. These new applications, coupled with the new features required for future networks, introduce changes in their structure, as well as in the way in which nodes may interact.

The most representative example of this situation is, perhaps, that of cellular communications, which have experienced and will experience massive traffic increases in the next years. This increase began with the shift from pure voice traffic to data traffic, and will be fueled mostly by smartphones, tablets and video streaming [1]. Both the number of devices and the data rate will grow exponentially, and new networks will have to cope with different requirements that each type of device may have, for example, in terms of data rate, latency and reliability [22]. This presents substantial challenges to communication engineers, which require novel and non-incremental advances, optimizing the way in which scarce resources, such as the wireless spectrum, are employed. This has ushered a new paradigm which departs from the original homogeneous single-tier company-planned network of sparse high-power base stations (BSs) to which mobile users connect. Future mobile networks will be multi-tier heterogeneous networks composed of BSs with different coverage areas and less centralized planning, in which new, more complex interactions will be possible, such as relaying, and inter-user cooperation. In this context, communication models which consider the characteristics of the wireless medium and the large, heterogeneous network premise are required in order to make these changes a reality.

3.2 Capacity notions in slow fading channels

One of the main goals in the study of communication systems is the study of the fundamental limits at which information can be *reliably* transmitted through a communications channel. A rigorous mathematical formulation of a communications channel, the notion of reliable communications and the first analyses of such limits were pioneered by C. E. Shannon in 1948 [5], leading to the creation of Information Theory [4, 23] which is today one of the most important and fundamental branches of human knowledge.

Shannon's first results addressed the point-to-point channel, that is, between one transmitter and its receiver, whose diagram can be seen in Fig. 3.1. The basic elements that define the channel are[1]:

- An alphabet $\mathcal{X}$ whose elements, known as symbols, can be input to the channel.
- An alphabet of output symbols $\mathcal{Y}$, which are the set of symbols which can be observed at the output of the channel.
- A conditional probability distribution $p_{Y^n|X^n}$, which, given a channel input sequence $X^n \in \mathcal{X}^n$ of length n, will be used to generate a channel output sequence $Y^n \in \mathcal{Y}^n$ of length n.

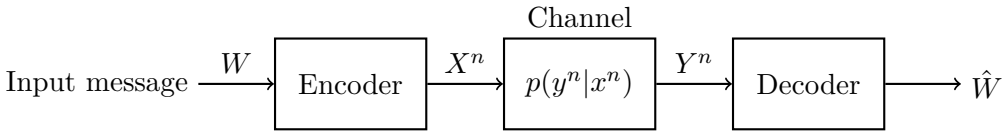

Figure 3.1: Block diagram of the mathematical model of a point-to-point channel.

A transmission of length n through the channel will require an (M, n) code, composed of the following:

- A set of messages $\{1, \ldots, M\}$ from which one, denoted by W, is chosen to be transmitted.
- At the transmitter: a coding function $X^n : \{1, \ldots, M\} \to \mathcal{X}^n$, which assigns a codeword of length n to each of the messages.
- At the receiver: a decoding function $Y^n : \mathcal{Y}^n \to \{1, \ldots, M\}$, which generates an estimate $\hat{W}$ of the transmitted message based on the output sequence.

[1] In this work, time is assumed to be discrete. In general, continuous-time channels can be converted to discrete-time, through an appropriate sampling procedure (see [4]).

The rate R of an (M, n) code (in bits/channel use) is defined as $R = \log_2(M)/n$, and a rate is said to be *achievable* if there exists a sequence of (M, n) codes such that the probability of incorrectly decoding a codeword goes to zero as the length of the code n goes to infinity. The capacity of the channel C is therefore defined as the supremum of all achievable rates.

One of the most frequently used models for communication channels is the additive white Gaussian noise channel (AWGN channel, Fig. 3.2) whose input-output relation is:

$$Y_i = gX_i + Z_i, \quad i = 1, 2, \ldots \tag{3.1}$$

with $\{Z_i\}_i$ a sequence of zero-mean independent complex circularly symmetric Gaussian (CCSG) random variables [6] with variance σ^2, independent of $\{X_i\}_i$, and g is a complex channel gain. Since the inputs are arbitrary complex numbers, if a transmission is carried

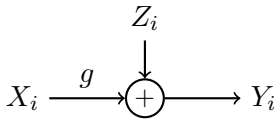

Figure 3.2: Block diagram of an additive white Gaussian noise channel given by (3.1).

out through this channel using an (M, n) code, the input symbols could be selected as spread out as desired for any rate, meaning the capacity of this channel would be infinite. For this reason, additional restrictions on the codes are considered. One of the most frequent ones is an energy restriction on the input sequence $(x_1, \ldots, x_n)$, as follows:

$$\frac{1}{n}\sum_{i=1}^{n} |x_i|^2 \leq P, \tag{3.2}$$

where $P > 0$ represents the constraint. Shannon showed that the capacity of this channel is:

$$C = \sup_{X:\mathbb{E}[|X|^2]<P} I(X, Y) = \log_2(1 + \text{SNR}), \tag{3.3}$$

where $\text{SNR} = |g|^2 P/\sigma^2$ is the signal-to-noise ratio (SNR) at the receiver. $I(X, Y)$ is the mutual information between the random variables X and Y, defined as:

$$I(X, Y) = \mathbb{E}\left[\log_2\left(\frac{f_{X,Y}(X, Y)}{f_X(X) f_Y(Y)}\right)\right], \tag{3.4}$$

where $f_{X,Y}(x, y)$, $f_X(x)$, $f_Y(y)$ are the density functions of the corresponding random variables. The supremum in (3.3) is with respect to all the probability distributions satisfying the indicated constraint on the variance, which can be attained by taking X as a zero-mean CCSG random variable of variance P. The proof also shows that this can be achieved by generating the codewords as independent CCSG random variables of variance P; although these

codes are not practical because of their decoding complexity, they are useful for analyzing the theoretical performance of communications systems.

The previous capacity result is valid when the channel coefficient g is fixed. However, in many cases this coefficient will have some dynamics, such as being time-varying, which implies that the capacity of the resulting channel will be different or, in some cases, that the standard definition of capacity may not be well-suited for it [24]. One of theses cases, the one we will consider in this work, is that of *slow-fading* channels. In simple terms, a channel is slow-fading when its time changes are *slow* compared to the duration of a transmission[2]. Therefore, in each transmission, a transmitter/receiver pair will observe a single realization of the channel. Since, in general, this realization could be arbitrarily bad, there is no positive rate at which transmission could reliably take place, and hence the traditional notion of capacity will not work. An outage analysis, on the other hand, establishes that an outage event will occur whenever the realization of the channel is *bad enough* such that the rate attempted by the user cannot be transmitted through the channel. The probability of this event or *outage probability* (OP) is therefore an important service metric for any slow fading channel.

To illustrate this, let us consider again the Gaussian channel (3.1) under slow fading[3], that is, the channel coefficient g is randomly chosen before the transmission starts according to a distribution F_g and is fixed for the duration of the transmission. This means that the capacity of the channel in the traditional sense will be a random variable tied to the fading coefficient:

$$C(g) = \log_2\left(1 + \frac{|g|^2 P}{\sigma^2}\right). \tag{3.5}$$

For each realization of the channel coefficient, the capacity will be given by (3.5). Since the channel coefficient could be, in general, arbitrarily close to zero, for any rate $R > 0$ we have a positive probability of being below this magnitude and hence the capacity is zero. In an outage analysis, the decoding procedure at the destination is examined, to determine the errors that will be committed in those *bad* realizations of the channel, and the probability of these events is found. In this example, the outage event for an attempted rate R is simply

[2]While transmissions in the real world are of finite duration, the notion of reliable communication introduced in the context of information theory considers transmissions which are arbitrarily long, so that the asymptotic behavior of the system can be considered. This means that, for the information theoretic analysis, the fading will remain constant for any transmission length, and hence, it will not be *slow*, as the name indicates, but fixed. The performance of a finite-length transmission will, in many cases, be very close to the asymptotic behavior without requiring an excessive length, and hence, the validity and importance of the asymptotic analysis.

[3]Actually this channel is also narrow band, which means that its effect on the input signal is represented by a scaling factor. See Appendix A for a brief introduction to the motivation behind the wireless models used in this work.

$\{C(g) < R\}$, and the OP is:

$$P_{\text{out}}(R) \triangleq \mathbb{P}(C(g) < R) = \mathbb{P}(|g|^2 < \sigma^2(2^R - 1)/P) = F_{|g|^2}\left(\frac{\sigma^2}{P}(2^R - 1)\right), \tag{3.6}$$

where $F_{|g|^2}$ is the cumulative distribution function of the power fading coefficient.

3.3 Stochastic geometry models of wireless networks

Later works on information theory extended Shannon's initial results considering more complex scenarios, including channels with more than one transmitter and/or receiver, such as the multiple-access, broadcast and interference channels [4, 23]. For several cases the traditional capacity is known, while for others only inner and outer bounds to the capacity regions are available. Although the analysis of networks with many nodes is in theory possible from an information-theoretic perspective, the expressions obtained are sometimes very complex and hard to interpret or grasp beyond the situation in which there are a few nodes. Wireless networks typically consist of many users, so suitable models, which lead to general and insightful conclusions, are needed. The conditions for signal propagation in a wireless medium are harsher than in most wired scenarios. Signals suffer from attenuation due the traveled distance, as well as from the (unpredictable) geometry of the space in which the transmissions take place. In typical wireless systems there will be many simultaneous transmissions taking place in space, and, due to the broadcast nature of the wireless medium and the limited spectrum available, interference between different users in the network will be present. In many cases, this interference will be larger and more detrimental to communications than the presence of background noise.

Let us consider a frequency non-selective slow fading wireless network with l transmitters and m receivers sharing the same space and spectrum. In each channel use, this network could be represented by a system of equations:

$$y_k = \sum_{j=1}^{n} a_{j,k} x_j + z_k, \qquad k = 1, \ldots, m \tag{3.7}$$

where y_k is the symbol received by receiver k, $a_{j,k}$ is a complex channel gain, x_j is the symbol transmitted by the user j, and z_k is Gaussian noise at receiver k. Some transmitters could have one or more messages to transmit to one or more users, while some transmitters may cooperate by acting as relays. Additionally, some users may cause interference to each other. Wireless models require considering a large number of users (large m and n), whose positions are unpredictable and influence their channel qualities and the interference. Hence, analyzing a specific spatial configuration may not yield insights to the average behavior of the users in

the network, and in general will not be possible in closed form, mainly because the asymmetry in the channels of a specific realization will lead to different performance levels for each node. Furthermore, the behavior of the interference is different to that of background noise, in the sense that the interference is affected by decisions that nodes take, i.e., the same selfish decision taken by all the users may worsen the performance of all the users. This does not happen when background noise, even correlated, is present at the receivers.

The theory of point processes is a branch of stochastic geometry [2] which studies random point patterns in a suitable space. Hence, it introduces a natural way of modeling and analyzing large wireless networks in which the users are spatially distributed in an unpredictable manner. Many of the aspects mentioned above can be considered, and averaging over all the possible realizations of the network, the average effects of, among other things, interference and its time-space correlation can be characterized through compact and closed form expressions. A point process is analog to a random variable except that the outcome of a realization is a set of points instead of a real number. In the context of wireless networks, points will be generally distributed on the plane, that is, on $\mathbb{R}^2$. The most common and widely used is the Poisson point process (PPP), which we introduce in Chapter 4. In the context of the AWGN channel, we have seen that the capacity is strongly defined by the SNR, while in the case in which interference is present, the parameter of interest will be the signal-to-inference-noise ratio (SINR) defined as the quotient between useful signal power and the power of the interference generated by the other users and the background noise. To see how this would be defined, let us consider a network like the one shown in Fig. 3.3, composed of many transmitter-receiver pairs distributed in space. Let us assume that transmissions are independent among the users, that no effort is made in decoding the interference, that the users employ unit power and that background noise of variance σ^2 is present at the receivers. We focus on a transmitter receiver pair and assume that the other transmitters in the network are distributed as a point process $\Phi = \{x_i\}_{i\in\mathbb{N}}$. In this case, the SINR at the destination d receiving a transmission from the source at s would be:

$$\mathrm{SINR}(s,d) = \frac{|g(s,d)|^2}{\sum_{x\in\Phi} |g(x,d)|^2 + \sigma^2}, \tag{3.8}$$

where $|g(x,y)|^2$ is the power fading coefficient from x to y. Since the interference is treated as noise and is uncorrelated among users, their individual powers add together. In many cases, the interference power will be much larger than the noise, so many times the signal-to-interference ratio (SIR) is considered.

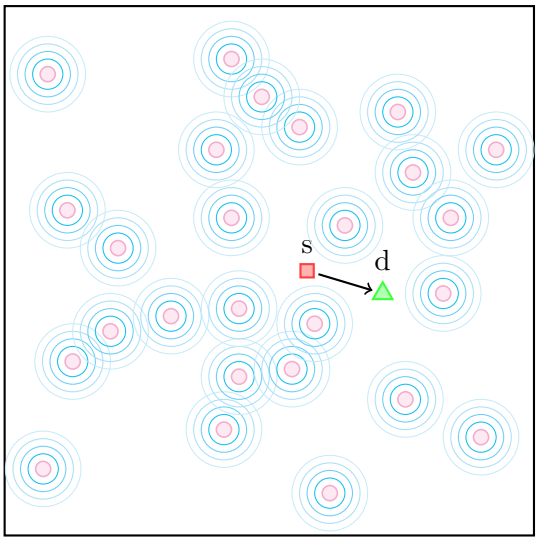

Figure 3.3: Wireless network with a reference source (square) transmitting to its destination (triangle) while other transmitters (circles) attempt to transmit to their destinations causing interference.

3.4 Modeling cooperative networks through stochastic geometry

Cooperative communications in wireless networks [11, 12], in which one or more users act as a distributed antenna array to improve the rate and reliability of their links, emerged several years ago as a means of coping with the increasing traffic demands in wireless networks. In general, considering a single cooperative wireless link with uncorrelated background noise at the receivers shows that cooperation will provide substantial improvements. Wireless links are, in many cases, not separated enough to be considered in isolation and, therefore, it is useful to consider the effect of neighboring transmitters. In addition, if users cooperate by transmitting or retransmitting other user's messages without control or coordination, interference levels may grow to a point in which the benefits of cooperation are no longer substantial enough. The main goal of this work is to study some cooperative schemes in wireless networks from the perspective of stochastic geometry, thus considering the adverse effects of the spatial interaction of the users through interference. In the following section we give some preliminaries and background on the chapters of this thesis, highlighting the main contributions.

3.4.1 On relaying in large wireless networks

Perhaps one of the earliest cooperative channels introduced is the relay channel [13], in which a transmitter communicates to a destination with the aid of an additional node, known as the relay (see Fig. 3.4). The main communication strategies for the relay channel were introduced

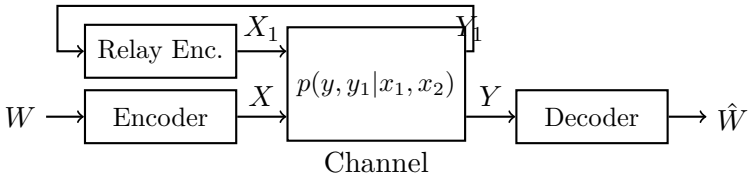

Figure 3.4: Representation of the mathematical model of the relay channel.

in [14]: decode-and-forward (DF) and compress-and-forward (CF). In the first one, the relay fully decodes the message sent by the source, re-encodes it and forwards it to the destination, which attempts to decode the message using both the transmission from the source and the relay. In compress-and-forward, the relay does not decode the message; instead, it finds a description from a known library which best represents the received signal, and forwards the index of this sequence to the destination. The destination also has access to the library and uses the received index to aid in decoding the transmission received from the source. This index acts as a compressed description of the received signal at the relay. Since in DF the relay has to decode the message of the source in order to transmit, this protocol will perform better when the source-relay channel is better than the source-destination channel, which in a spatial scenario will mean that the relay is not too far from the source. On the other hand, since CF does not require the relay to decode the message, it will work better when the relay-destination channel is good enough to sustain the transmission of the compressed version of the received sequence, when the relay is on average closer to the destination in a spatial scenario. Considering a single relay-channel in isolation with the impairments of uncorrelated background noise at the receivers has shown to yield interesting gains in communications quality with respect to point-to-point links [15, 16]. In Chapters 5 and 6, we study the performance of these protocols in the context of spatial models.

In Chapter 5, we first study the performance, in terms of the OP, of a single full-duplex relay channel when the interference is generated by a set of uniformly distributed interfering nodes, modeled by a PPP. This analysis serves as a starting point for studying how traditional coding schemes, which have been developed and studied in the context of classical information theory [4, 23], should be considered in the context of outage-based point process models. This is different to usual scenarios, in which interference comes from uncorrelated background

noises; the interference time signals at the destination and at the relay are correlated because they are generated by the same set of interfering nodes, and, at the same time, this correlation becomes random when the random distribution of nodes is considered through the point process. In the case of DF, the achievable rate does not depend on the spatial correlation of the interference time signals, because of the requirement that the relay fully decode the message of the source. For this reason, a closed form expression to the outage probability of DF can be found. In the case of CF, the achievable rates depend on this correlation because the relay does not fully decode the message, and the characterization of the probability distribution of this correlation is not possible. Therefore, we provide an upper bound to the outage probability of this channel by means of constant gaps on the achievable rates. We compare the outage probability performance of these two protocols with a point-to-point transmission (DT) and with a half-duplex DF protocol.

Studying a single relay channel in the presence of interference represents a best case from the perspective of the user who has the relay. However, if more users in the network use one, it will also increase the interference levels at other receivers, thus reducing the benefits of cooperation. A balance is introduced between the benefits of cooperation for a user and the increased levels of interference it generates throughout the network. This may reduce or altogether cancel the benefits of cooperation that are observed when a single user with a relay is considered. This motivates the results in Chapter 6, in which we consider a decentralized wireless network, and analyze the case in which the interfering users may also use a relay. In particular, we focus on the full-duplex DF protocol and consider that each transmitter in the network may use a relay, which is chosen as its nearest neighbor from the set of nodes which are inactive. In this setup, we introduce a simple activation strategy, in which the relays chose to activate themselves in a random independent manner. We focus in the regime in which the OP in the network is in the practical operating range, and study its behavior under different relay activation probabilities, finding the values which minimize the OP.

3.4.2 On device-to-device communications in large wireless networks

After studying the use of relays in spatial models, we then focus on cooperation from the perspective of content-sharing between users in cellular networks. As we mentioned before, traffic in cellular networks has experienced a large increase since the transition from voice only communications (up to 2G) to a majority of data traffic (starting from 3G onwards). The most effective way to improve the rate and reliability of cellular networks has been found to be a reduction in cell sizes and transmission distances [17], that is, an increase in the spatial frequency reuse, and spectral efficiency. One of the possible strategies that aim in this direction is that of device-to-device (D2D) communications [18], in which mobile users in close

proximity may establish a direct link through a short direct-hop instead of connecting to the BS [25]. This will increase the frequency reuse, energy efficiency and latency, among other things. Predicted increases in mobile data traffic will be motivated largely by an increase in on-demand video requests [1], in which a large number of users access a small number of files, in an asynchronous fashion, such that broadcast transmissions cannot be employed [20]. In this setup, the unused storage space at the users may be used to cache watched video files, which could then be exchanged through D2D communications, thus avoiding a downlink file transfer from the BSs. In addition, if exchanges take place outside the downlink frequency (*out-of-band* D2D), the spectrum reuse is increased, freeing up resources for other users wanting to use the cellular frequency bands. In Chapter 7, we approach this problem through a point process model. More specifically, we are interested in studying the fraction of user video requests that could be served in this way, thus avoiding the downlink transmission. This fraction is an indication of the potential economical and quality-of-service benefits that could be achieved through this strategy. This problem cannot be analyzed only on the link layer, that is, by only studying if a transmission fails or not. This is because sharing users will have a requested file with a certain probability, and this has to be considered to determine whether a request may be served through D2D or not. In fact, the problem has three main aspects:

- The file caching policy for the users, that is, which files the users chose to store, how many, etc.
- The problem of user matching for file exchanges, that is, who should share what with whom.
- Problems associated with transmissions, such as, scheduling (who should go first) and link quality.

One of the main contributions in this chapter is introducing a spatial framework in which these aspects are considered. We introduce a model in which users are grouped into spatial clusters, where file exchanges take place, and define trade-off regions which pertain the average fraction of video requests that could be served through D2D in time-block, both globally and per cluster, when a quality constraint is imposed on the transmission links. In addition, we also introduce a trade-off region that relates the global and local (per cluster) fraction of served requests. Using this framework, we consider a simple file exchanging protocol in which time-division access is used inside each cluster. By analyzing the time-correlated interference during the time-block, we evaluate the metrics for this simple strategy, which give inner-bounds to the trade-off regions, and hence to the fraction of requests that could be served. The ideas used in developing this framework may be used in future works to develop simple

cross-layer analysis of networks, thus allowing more macroscopic network characteristics to be considered.

Chapter 4

Relevant Results on Point Processes

Summary. This chapter gives an overview on certain aspects of the theory of point processes, and the relevant results that will be required in the next chapters. The reader who is familiar with these results can skip this chapter and, if needed, come back to a specific result when it is referenced in the text.

4.1 Introduction

In many real world applications, the outcome of an experiment or the behavior of a system can be modeled through a point pattern occurring in a suitable space. When the position of these points is unpredictable, modeling these distributions as a random point pattern is a good alternative. The theory of point processes provides a rigorous and powerful mathematical framework and toolbox for defining and analyzing such patterns. In our case, for example, we will be mostly interested in using a point process on $\mathbb{R}^2$ to model the unpredictable distribution of users in a large wireless network.

In this section we will provide a brief review on the definition of a point process and the main tools that we will require in the next chapters. For readers wanting more details, good starting points for the study of point processes are [26] and the highly referenced [2]. For a rigorous and general treatment on the topic the reader is directed to [27, 28, 29]. For an introduction to point processes and their usage in wireless communications the reader is directed to [7, 8, 9, 30, 31]. Finally, a specific text on Poisson processes is [32].

4.2 Main definitions

This section provides a brief overview of what a point process is. Some aspects of the construction of a point process will be mentioned, without rigor, for the sake of better understanding their mathematical structure.

To define the point process we require a *state space*, that is, the space where the points will lie. In our case, for modeling the positions of nodes, this will be the measurable space $(\mathbb{R}^2, \mathcal{B}_{\mathbb{R}^2})$, where $\mathcal{B}_{\mathbb{R}^2}$ is the usual Borel σ-algebra. To give the general definition of a point process, which will be useful for defining marked point processes, we shall consider the state space to be a complete separable metric space $\mathcal{X}$, with a σ-algebra $\mathcal{B}_{\mathcal{X}}$ of Borel sets induced by the metric defined on the space [28].

Definition 4.2.1 (Measure)**.** *Given a measurable space $(\mathcal{X}, \mathcal{A})$, a function ξ on $\mathcal{A}$ is a measure if it is non-negative and given a family of sets $\{A_i\}_{i\in\mathbb{N}} \subseteq \mathcal{A}$ mutually disjoint, then:*

$$\xi\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \sum_{i\in\mathbb{N}} \xi(A_i). \tag{4.1}$$

One of the most simple examples of measures are the Dirac measures, which are defined for any $x \in \mathcal{X}$ as:

$$\delta_x(A) = \mathbb{1}\{x \in A\}, \quad A \in \mathcal{A}. \tag{4.2}$$

To define the point process, we will use measure space $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$, and define:

Definition 4.2.2 (Counting measure)**.** *Given a measurable space $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$, a* counting measure *is a function $\phi : \mathcal{B}_{\mathcal{X}} \mapsto \mathbb{N}_0 \cup \{\infty\}$ such that ϕ is a measure and ϕ is boundedly finite, that is, for every bounded $A \in \mathcal{B}_{\mathcal{X}}$, $\phi(A) < \infty$.*

For each counting measure ϕ there exists a family of points $\{x_i\}_i \subseteq \mathcal{X}$, and a family of non-negative integer weights $\{a_i\}_i$ such that ϕ is expressed as a superposition of counting measures:

$$\phi = \sum_i a_i \delta_{x_i}, \tag{4.3}$$

implying that ϕ will measure every $A \in \mathcal{B}_{\mathcal{X}}$ as:

$$\phi(A) = \sum_i a_i \delta_{x_i}(A). \tag{4.4}$$

A counting measure will be *simple* when $a_i = 1$ for all i. It is straightforward to observe that a counting measure can be used to represent a point pattern, such as the spatial distribution of nodes in a network, by representing the positions of the points by the family $\{x_i\}_i$ of weighted points of the measure. The fact that a counting measure is boundedly finite implies that a finite region in space will have a finite number of points, and a simple counting measure guarantees that points will not be superimposed. In addition, these conditions imply that the number of points is countable, that is, the natural number can be used as an index set for the family, and the sum in (4.4) will have a countable number of terms. Finally, given a set $A \in \mathcal{A}$ taking $\phi(A)$ will give the number of nodes of the network which are in A.

In order to represent a random pattern we now need to define a random element, a generalization of a random variable, which, for every realization will give a realization of a counting measure/point pattern. For this we need to introduce the following:

Definition 4.2.3. *We denote by $\mathcal{N}_{\mathcal{X}}^{\#}$ the set of all counting measures over the state space $\mathcal{X}$. This is the set of all numerable point patterns which have a finite number of points in bounded sets.*

A topology, known as the weak-hash topology, can be defined in $\mathcal{N}_{\mathcal{X}}^{\#}$ [28]. This topology defines the family of open sets in $\mathcal{N}_{\mathcal{X}}^{\#}$ and allows the definition of the σ-algebra generated by these open sets or Borel σ-algebra, denoted as $\mathcal{B}_{\mathcal{N}_{\mathcal{X}}^{\#}}$.

Definition 4.2.4 (Point process)**.** *A point process Φ on $\mathcal{X}$ is a measurable mapping from a probability space $(\Omega, \Sigma, \mathbb{P})$ to $(\mathcal{N}_{\mathcal{X}}^{\#}, \mathcal{B}_{\mathcal{N}_{\mathcal{X}}^{\#}})$, that is, a mapping:*

$$\Phi : \Omega \to \mathcal{N}_{\mathcal{X}}^{\#}, \tag{4.5}$$

such that for every $A \in \mathcal{B}_{\mathcal{N}_{\mathcal{X}}^{\#}}$, we have that $\Phi^{-1}(A) \in \Sigma$.

A point process is like a usual random variable with the exception that a realization is not a real number, but a counting measure. We may denote each realization as $\Phi(\omega)$ or as $\Phi(\cdot, \omega)$, where the first component indicates that each realization of Φ can be applied to any $A \in \mathcal{B}_{\mathcal{X}}$ to count the number of points in A, like in (4.4). In fact, if we fix $A \in \mathcal{B}_{\mathcal{X}}$ we could interpret $\Phi(A, \cdot) : \Omega \to \mathbb{N}_0 \cup \{\infty\}$ as a map which, being defined from the probability space $(\Omega, \Sigma, \mathbb{P})$, is a candidate to be a random variable. For short, we can denote this function as $\Phi(A)$.

Lemma 4.1 ([28, Corollary 9.1.IX])**.** *A map $\Phi : \Omega \to \mathcal{N}_{\mathcal{X}}^{\#}$, will be a point process if and only if $\Phi(A)$ is a random variable for every bounded $A \in \mathcal{B}_{\mathcal{X}}$.*

This means that a point process can also be interpreted as a family of random variables $\{\Phi(A)\}_{A \in \mathcal{B}_{\mathcal{X}}}$. In fact, the σ-algebra $\mathcal{B}_{\mathcal{N}_{\mathcal{X}}^{\#}}$ can also be shown to be the smallest one such that all the mappings $\{\Phi(A)\}_{A \in \mathcal{B}_{\mathcal{X}}}$ are measurable.

A point process is said to be *simple* when all its realizations are almost surely simple, that is:

$$\mathbb{P}(\{\omega : \Phi(\omega) \text{ is simple}\}) \equiv \mathbb{P}(\Phi \text{ is simple}) = 1. \tag{4.6}$$

In addition, the probability measure $\mathbb{P}$ induces, through the process Φ a new measure $\mathbf{P}$ on the measurable space $(\mathcal{N}_{\mathcal{X}}^{\#}, \mathcal{B}_{\mathcal{N}_{\mathcal{X}}^{\#}})$, in the following way:

$$\mathbf{P}(A) = \mathbb{P}(\{\omega : \Phi(\omega) \in A\}) = \mathbb{P}(\Phi \in A), \quad A \in \mathcal{B}_{\mathcal{N}_{\mathcal{X}}^{\#}}. \tag{4.7}$$

The measure $\mathbf{P}$ is known as the distribution of the point process Φ.

Definition 4.2.5 (Stationary point process)**.** *A point process Φ is stationary if its distribution is invariant under translations, that is, for any $A \in \mathcal{B}_{\mathcal{N}_{\mathcal{X}}^{\#}}$ and any fixed $v \in \mathcal{X}$:*

$$\mathbb{P}(\Phi \in A) = \mathbb{P}(\Phi_v \in A), \tag{4.8}$$

where $\Phi_v = \{x + v : x \in \Phi\}$ is the displaced point process. This can also be written as: $\mathbf{P}(A) = \mathbf{P}(A_{-v})$, where $A_v = \{\phi \in \mathcal{N}_{\mathcal{X}}^{\#} : \phi_{-v} \in A\}$.

Although we have defined what a point process is, it is not clear so far how to construct one from this definition. For this, it is useful to introduce the finite-dimensional distributions of the point process:

Definition 4.2.6 (Finite-dimensional distributions)**.** *The finite-dimensional distributions of a point process are the joint distributions, for all the finite families of bounded Borel sets $A_1, \ldots, A_k$, of the random variables $\Phi(A_1), \ldots, \Phi(A_k)$.*

We will not go into all the details here, but it is enough for our purposes to say that it is necessary and sufficient to specify the set of finite-dimensional distributions in order to fully and uniquely specify a point process [28, Theorem 9.2.X].

With this, we are now able to introduce a specific point process, which will be used in the rest of this work:

Definition 4.2.7 (Poisson process)**.** *Given the space $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$ and a locally-finite non-atomic*[1] *measure on Λ on $\mathcal{X}$, a point process Φ on $\mathcal{X}$ is a Poisson point process (PPP) if:*

- *The number of points of Φ in any bounded $A \in \mathcal{B}_{\mathcal{X}}$ is a Poisson random variable of mean $\Lambda(A)$.*
- *If $A_1, \ldots, A_n$ are disjoint bounded sets in $\mathcal{B}_{\mathcal{X}}$ then the number of points are independent random variables.*

If $\mathcal{X} = \mathbb{R}^d$, then a Poisson process is stationary or *homogeneous* if Λ is a multiple of the Lebesgue measure, that is $\Lambda(A) = \lambda|A|$ where $|A|$ is the area of A, and $\lambda > 0$ is a constant.

Definition 4.2.8 (Mean measure)**.** *When a simple point process has finite average number of points in any bounded Borel set, its intensity measure or mean measure is said to exist, and is defined as*

$$\Lambda(A) \triangleq \mathbb{E}\left[\Phi(A)\right], \qquad A \in \mathcal{B}_{\mathcal{X}}. \tag{4.9}$$

[1] A measure Λ is non-atomic if there is no $x \in \mathcal{X}$ such that $\Lambda(x) > 0$.

If Φ is a simple stationary point process, its mean measure has to be translation-invariant; in addition if its state space is $\mathcal{X} = \mathbb{R}^d$, any measure which is translation invariant has to be a multiple of the Lebesgue measure. Therefore, for every stationary simple point process Φ on $\mathbb{R}^d$ we can find $\nu > 0$ such that:

$$\Lambda(A) = \nu|A|, \tag{4.10}$$

where ν is called the *intensity* of the process and has units of nodes/area. In the case of the homogeneous PPP, the constant λ is its intensity.

In addition to its finite-dimensional distributions, a point process may also be characterized through its Laplace functional:

Definition 4.2.9 (Laplace functional)**.** *The Laplace functional of a point process Φ is defined as:*

$$\mathcal{L}_\Phi(f) = \mathbb{E}\left[e^{-\sum_{x\in\Phi} f(x)}\right] \tag{4.11}$$

for any non-negative measurable function f.

Notice that this expectation involves averaging with respect to the distribution of the point process, which is hard to do in general. For the Poisson process this can be done, as the following Lemma states:

Lemma 4.2 (Laplace functional of a Poisson process)**.** *The Laplace functional of a Poisson point process Φ of intensity measure Λ is:*

$$\mathcal{L}_\Phi(f) = \exp\left\{-\int_{\mathcal{X}} \left(1 - e^{f(x)}\right) \Lambda(dx)\right\}. \tag{4.12}$$

The proof of this can be found in [8].

4.3 Marked point processes

In most cases, the distribution of nodes in a network will be modeled as a homogeneous PPP in $\mathbb{R}^2$, and we will denote this process as $\Phi = \{x_1, x_2, \ldots\} = \{x_i\}$. In some cases, we will be interested in associating certain information to a node, for example, the fading coefficient of a transmitter towards its destination. This attached information will be a *marking* of the point. For this purpose, the notion of marked point process is useful. In simple terms, a marked point process can be formed by starting from a point process $\Phi = \{x_i\}$ and attaching a random vector $\mathbf{m}_{x_i}$ to each point, thus forming a new family of points $\tilde{\Phi} = \{(x_i, \mathbf{m}_{x_i})\}$ in a larger space. Formally, we will define a marked point process as a point process $\tilde{\Phi} = \{(x_i, \mathbf{m}_{x_i})\}$ such that the original or *ground* process $\Phi = \{x_i\}$ is also a point process.

Definition 4.3.1 (Marked point process)**.** *A marked point process on a space $\mathcal{X}$ with marks in a space $\mathcal{M}$ is a point process $\tilde{\Phi}$ on $\mathcal{X}\times\mathcal{M}$ such that the ground measure $\Phi(A) \triangleq \tilde{\Phi}(A\times\mathcal{M})$ is in $\mathcal{N}_{\mathcal{X}}^{\#}$, that is, the points without their marks form a point process themselves.*

Definition 4.3.2 (Independently marked point process)**.** *Given a marked point process $\tilde{\Phi} = \{(x_i, \mathbf{m}_{x_i})\}$ on $\mathcal{X}\times\mathcal{M}$, we say it is independently marked if, given the ground process $\tilde{\Phi}$, the marks $\{\mathbf{m}_{x_i}\}$ are independent random variables, whose distribution depends only on x.*

The probability structure of $\tilde{\Phi}$ is completely specified by the distribution of the ground process Φ and a mark kernel $\{F_x(A) : A \in \mathcal{B}_{\mathcal{M}}, x \in \mathcal{X}\}$ representing the conditional distribution of the mark, given the location x.

Definition 4.3.3 (Stationary marked point process)**.** *A marked point process $\tilde{\Phi} = \{(x_i, \mathbf{m}_{x_i})\}$ on $\mathcal{X}\times\mathcal{M}$ is stationary if its distribution is invariant under translations of its first component, that is, on translations of the points and not of the marks:*

$$(x, \mathbf{m}_x) \mapsto (x + v, \mathbf{m}_x).$$

Given a marked point process, we are also interested in finding its Laplace functional, which in general could be very hard. Fortunately, for the independently marked PPP it is known in closed form, as the following lemma states:

Lemma 4.3. *An independently marked process $\tilde{\Phi} = \{x_i, \mathbf{m}_{x_i}\}$ formed by a Poisson process Φ with intensity measure Λ and marks in $\mathcal{M}$ with distribution $F_x(dm)$, is a Poisson point process in the space $\mathcal{X}\times\mathcal{M}$ with intensity measure:*

$$\tilde{\Lambda}(A\times K) = \int_A \int_K F_x(dm)\Lambda(dx), \quad A \in \mathcal{B}_{\mathcal{X}}, K \in \mathcal{B}_{\mathcal{M}}. \tag{4.13}$$

This implies that the Laplace functional of an independently marked Poisson process is:

$$\mathcal{L}_{\tilde{\Phi}}(f) = \mathbb{E}\left[e^{-\sum_{x\in\Phi} f(x,\mathbf{m}_x)}\right] = \exp\left\{-\int_{\mathcal{X}}\left(1 - \int_{\mathcal{M}} e^{-f(x,\mathbf{m})}F_x(d\mathbf{m})\right)\Lambda(dx)\right\}, \tag{4.14}$$

for $f : \mathcal{X}\times\mathcal{M} \to \mathbb{R}_{\geq 0}$. Finally, when $\mathcal{X} = \mathbb{R}^2$ and the process Φ is homogeneous then $\Lambda(dx) = \lambda dx$. If the process is stationary then $F_x(d\mathbf{m}) \equiv F(d\mathbf{m})$.

The proof of this lemma can be found in [8, 32].

4.4 Averages, conditioning and Palm distributions

Now consider a point process Φ with intensity function Λ. One may be interested in studying the average of random sums over the point process:

$$\mathbb{E}\left[\sum_{x\in\Phi} f(x)\right], \tag{4.15}$$

where f is a non-negative measurable function. Choosing f appropriately, this could be, for example, the average interference power that is observed at a fixed point in space.

Lemma 4.4 (Campbell's Formula)**.** *Let Φ be a point process with intensity measure Λ, and $f : \mathcal{X} \to \mathbb{R}$ a measurable function. Then we have:*

$$\mathbb{E}\left[\sum_{x\in\Phi} f(x)\right] = \int_{\mathcal{X}} f(x)\Lambda(dx). \tag{4.16}$$

This formula allows us to write the expectation with respect to the point process distribution as an integral in $\mathcal{X}$ with respect to the mean measure. In the case in which $\mathcal{X} = \mathbb{R}^2$ and the process is stationary, this is a simple integration in $\mathbb{R}^2$.

A more complex scenario, yet more interesting, would be to consider functions which depend on the whole point process. For this, it is convenient to introduce the Campbell measure:

Definition 4.4.1 (Campbell measure)**.** *For a set $\tilde{A} \in \mathcal{B}_{\mathcal{X}} \otimes \mathcal{B}_{\mathcal{N}_{\mathcal{X}}^{\#}}$, where $\otimes$ denotes the product σ-algebra, we define the Campbell measure of $\tilde{A}$ as:*

$$\mathbf{C}(\tilde{A}) \triangleq \mathbb{E}\left[\sum_{x\in\Phi} \mathbb{1}\{(x,\Phi) \in \tilde{A}\}\right]. \tag{4.17}$$

On a product set, $\tilde{A} = A \times K$, this measure has a simple interpretation:

$$\mathbf{C}(A \times K) = \mathbb{E}\left[\Phi(A)\mathbb{1}\{\Phi \in K\}\right]. \tag{4.18}$$

If the intensity measure Λ exists and is boundedly finite, the application of the Radon-Nikodym theorem, indicates that the Campbell Measure can be written as [28]:

$$\mathbf{C}(A \times K) = \int_A \mathbf{P}^x(K)\Lambda(dx), \tag{4.19}$$

where $\{\mathbf{P}^x\}_x$ is a kernel of Palm probability distributions associated to the point process Φ at the location x. These are interpreted as conditional distributions of the point process having a point at x. This leads to a generalization of the Campbell formula:

Theorem 4.5 (Campbell-Mecke formula)**.** *For any measurable function $f : \mathcal{X} \times \mathcal{N}_{\mathcal{X}}^{\#} \to \mathbb{R}_{\geq 0}$ which is integrable with respect to the Campbell measure, we have:*

$$\mathbb{E}\left[\sum_{x\in\Phi} f(x,\Phi)\right] = \int_{\mathcal{X}\times\mathcal{N}_{\mathcal{X}}^{\#}} f(x,\phi)\mathbf{C}(d(x,\phi)) = \int_{\mathcal{X}}\int_{\mathcal{N}_{\mathcal{X}}^{\#}} f(x,\phi)\mathbf{P}^x(d\phi)\Lambda(dx). \tag{4.20}$$

In an analogous fashion, a reduced or modified Campbell measure can be introduced, which has the same principle, but which removes the point at x before averaging. This leads to the definition of a reduced Palm kernel $\{\mathbf{P}^{x!}\}_x$, which is interpreted as the distribution of the point process as seen from a point at x but removing this point, that is:

$$\mathbf{P}^{x!}(A) \triangleq \mathbf{P}^x(\{\phi : \phi \setminus \{x\} \in A\}). \tag{4.21}$$

Finally, this leads to the reduced Campbell-Mecke formula:

$$\mathbb{E}\left[\sum_{x\in\Phi} f(x, \Phi \setminus \{x\})\right] = \int_{\mathcal{X}} \int_{\mathcal{N}_{\mathcal{X}}^{\#}} f(x,\phi)\mathbf{P}^{x!}(d\phi)\Lambda(dx). \tag{4.22}$$

When the process Φ is stationary, then the Palm distributions are equivalent under translations, that is, $\mathbf{P}^x(A) = \mathbf{P}^0(A_{-x})$. This means that we only need the Palm distribution of a point at the origin and a shift operator to characterize the Palm kernel. This allows the introduction of the notion of a *typical point* of the process, in the sense that any point will statistically be the same.

In some cases, it may be necessary to consider other random objects which are related but not a part of the point process itself. It may be interesting to consider the distribution of these objects conditional to the point process having a point somewhere. In order to do this, for the case of stationary point processes, it is possible to define a measure $\mathbb{P}^0$ on the canonical measurable space (Ω, Σ), called the Palm measure of the point process. This defines a new probability space $(\Omega, \Sigma, \mathbb{P}^0)$, under which the distribution of all the random objects that were previously defined can be determined. The distribution of the point process Φ under $\mathbb{P}^0$ will be $\mathbf{P}^0$, and denoting by $\mathbb{E}^0$ the expectation with respect to $\mathbb{P}^0$, the Campbell-Mecke formula can be written as [8, 29]:

$$\mathbb{E}\left[\sum_{x\in\Phi} f(x, \Phi)\right] = \int_{\mathcal{X}} \mathbb{E}^0\left[f(x, \Phi_{-x})\right] \Lambda(dx). \tag{4.23}$$

The Palm distributions are, in most cases, unknown; fortunately, this is not the case for the Poisson process:

Theorem 4.6 (Slivnyak-Mecke)**.** *Φ is a Poisson point process with boundedly finite intensity measure Λ if and only if its distribution satisfies for all $x \in \mathcal{X}$:*

$$\mathbf{P}(\cdot) = \mathbf{P}^{x!}(\cdot) \tag{4.24}$$

This means that if we want to study the distribution of a PPP as seen from one of its points, but without counting this point, it is sufficient to add a point anywhere and consider the distribution of the point process with all its points.

Finally, we will be interested in studying this in the context of independently marked point processes. As we mentioned before, a marked point process can be interpreted as a point process in a higher dimensional space, encompassing the marks. In particular, for the case of independently marked Poisson processes, this process in a higher dimensional space will also be a Poisson process. A marked point process in $\mathcal{X}$ with marks in $\mathcal{M}$ can therefore be interpreted as an element on $\mathcal{N}^{\#}_{\mathcal{X}\times\mathcal{M}}$ and the Campbell measure (Def. 4.4.1), defined on $\mathcal{B}_{\mathcal{X}\times\mathcal{M}} \otimes \mathcal{B}_{\mathcal{N}^{\#}_{\mathcal{X}\times\mathcal{M}}}$, would be:

$$\mathbf{C}(\tilde{A}) = \mathbb{E}\left[\sum_{(x,\mathbf{m})\in\tilde{\Phi}} \mathbb{1}\{(x,\mathbf{m}) \in \tilde{\Phi}\}\right]. \tag{4.25}$$

This leads to the following formulas [8, 29]:

Theorem 4.7 (Campbell-Mecke formula for marked point processes)**.** *Let $\tilde{\Phi}$ be an independently marked point process with marks in $\mathcal{M}$ and mean measure Λ. For any measurable function $f : \mathcal{X}\times\mathcal{M}\times\mathcal{N}^{\#}_{\mathcal{X}\times\mathcal{M}} \to \mathbb{R}_{\geq 0}$ which is integrable with respect to the Campbell measure, we have:*

$$\mathbb{E}\left[\sum_{(x,\mathbf{m})\in\tilde{\Phi}} f(x,\mathbf{m},\tilde{\Phi})\right] = \int_{\mathcal{X}}\int_{\mathcal{M}\times\mathcal{N}^{\#}_{\mathcal{X}\times\mathcal{M}}} f(x,\mathbf{m},\tilde{\phi})\mathbf{P}^x(d(\mathbf{m},\tilde{\phi}))\Lambda(dx). \tag{4.26}$$

where $\mathbf{P}^x$ is a probability distribution on $\mathcal{M}\times\mathcal{N}^{\#}_{\mathcal{X}\times\mathcal{M}}$ called the Palm distribution of the marked point process $\tilde{\Phi}$ with a point at x. We can interpret $\mathbf{P}^x(A\times\cdot)$, for $A\in\mathcal{B}_{\mathcal{M}}$, as the distribution of the point process Φ with a point at x and mark vector in A.

This distribution $\mathbf{P}^0$ is obtained by disintegrating the Campbell measure. This Palm distribution itself can be disintegrated with respect to the mark distribution to obtain the following formula:

$$\mathbb{E}\left[\sum_{(x,\mathbf{m})\in\tilde{\Phi}} f(x,\mathbf{m},\tilde{\Phi})\right] = \int_{\mathcal{X}}\int_{\mathcal{M}}\int_{\mathcal{N}^{\#}_{\mathcal{X}\times\mathcal{M}}} f(x,\mathbf{m},\tilde{\phi})\mathbf{P}^{(x,\mathbf{m})}(d\tilde{\phi})F_x(d\mathbf{m})\Lambda(dx). \tag{4.27}$$

The Palm distribution $\mathbf{P}^{(x,\mathbf{m})}$ is the distibution of $\tilde{\Phi}$ under the condition that the ground process has a point at x with mark vector $\mathbf{m}$. The same formulas are valid for the reduced Palm distribution.

For details and proofs see [8, 28, 29].

In the case of stationary, independently marked point processes we may also consider the marked point process under the Palm probability $\mathbb{P}^0$ of the ground process Φ. That is, instead of considering the marked process as a process in a space with a higher dimension (a ground process with marks) and consider the Palm distribution of this process, we may

consider a modified probability space $(\Omega, \Sigma, \mathbb{P}^0)$ in which both the ground process Φ and the marks are defined at the same time. Denoting as $\mathbb{E}^0$ the expectation with respect to $\mathbb{P}^0$, the Campbell-Mecke formula (4.26) can therefore be written as:

$$\mathbb{E}\left[\sum_{(x,\mathbf{m})\in\tilde{\Phi}} f(x,\mathbf{m},\tilde{\Phi}_{-x})\right] = \int_{\mathcal{X}} \mathbb{E}^0\left[f(x,\mathbf{m},\tilde{\Phi}_{-x})\right]\Lambda(dx), \tag{4.28}$$

where $\tilde{\Phi}_{-x}$ denotes the marked process with the ground process translated by $-x$.

Chapter 5

On the Outage Probability of the Full-Duplex Interference-Limited Relay Channel

Summary. In this chapter, we study the performance, in terms of the asymptotic error probability, of a user which communicates with a destination with the aid of a full-duplex in-band relay. We consider that the network is interference-limited, and interfering users are distributed as a PPP. In this case, the asymptotic error probability is upper bounded by the OP. We investigate the outage behavior for well-known cooperative schemes, namely, DF and CF considering fading and path loss. For DF we determine the exact OP and develop upper bounds which are tight in typical operating conditions. Also, we find the correlation coefficient between the source and relay signals which minimizes the OP when the density of interferers is small. For CF, the achievable rates are determined by the spatial correlation of the interferences, and a straightforward analysis isn't possible. To handle this issue, we show the rate with correlated noises is at most one bit worse than with uncorrelated noises, and thus find an upper bound on the performance of CF. These results are useful to evaluate the performance and to optimize relaying schemes in the context of full-duplex wireless networks.

5.1 Introduction

In recent years, diversity-exploiting techniques for cooperative communications in wireless networks have been one of the most promising techniques to cope with the always increasing traffic demands. As such, strategies involving relays have received much attention as a means for improving the throughput and reliability of individual links [15, 16, 33]. For example, in the context of cellular networks, fourth-generation (4G) mobile-broadband systems allow for new coordination and cooperation strategies among base stations and relaying nodes. Basic relaying functionality was included in the Long Term Evolution (LTE) Rel. 10 standard [34], while Rel. 11 introduces several Coordinated Multi-Point Operation (CoMP) modes [35]. In the former, communications can be established via a half-duplex relay node, wirelessly con-

nected to a BS, either in the same frequency as the relay-destination link (in-band relaying), or in another band (out-of-band relaying) [34]. The latter considers homogeneous, i.e. single-tier, networks, with relays, and heterogeneous ones, where macro-cells and smaller cells are jointly deployed [35].

Since the seminal paper by Cover and El Gamal [14], the relay channel has received much attention from an information-theoretic perspective. In [14], the main coding strategies, DF and CF, were introduced and analyzed. A third alternative is amplify-and-forward which we will not address here. In general, works on additive-noise relay channels assume that transmissions are impaired by uncorrelated background noises at the receivers. Although this assumption has proved to be very useful, in the context of wireless networks, it may also be interesting to consider the interference correlation, which arises, among other things, because many receivers will experience interference from the same sources. This correlation appears both in the interference time signals, and in the random, correlated interference powers at the receivers. Moreover, in a large network, users may interact, causing each other adverse interference conditions, an effect which is not present when uncorrelated background noises are considered.

In this chapter, we study the performance, in terms of the asymptotic error probability, of full-duplex in-band relaying in a network impaired by interference between neighboring nodes, using an information-theoretic and stochastic geometry approach. Specifically, we consider the case in which a source attempts to communicate to a destination using the full-duplex DF or CF protocols. The gains in asymptotic error probability are upper bounded in terms of the OP, that is, the probability that, due to instantaneous conditions, the channel cannot support the rate attempted by a transmitting user. The co-channel interference experienced by the users is modeled using a homogeneous PPP, considering both path loss attenuation and Rayleigh fading.

5.1.1 Related work

The effects of interference power correlation and its impact on network performance have been analyzed in a number of papers such as [36, 37]. In the context of outage in slow-fading Gaussian networks with full-duplex relaying, previous works generally consider that fixed-power, uncorrelated noises are present at the receivers, and perform analysis for each value of the noise powers, without considering correlated or random interference signals. For example, [38] studies, among other things, the performance of DF in a single relay network in which the source-relay link has a fixed, known amplitude, and the source does not know if the relay is present or not. The expected throughput of the scheme and the optimal correlation between the transmissions of the source and the relay are characterized for each

value of the noise powers. In [39], a single-relay model in a Rayleigh fading environment is considered, with fixed, uncorrelated Gaussian background noise. The OP and the ergodic rate for DF and the correlation coefficient between the source and the relay, along with power allocation between transmitters are considered. Some examples in which correlation is considered between the nodes are [40] and [41]. Reference [40] studies the impact of noise correlation on the achievable rates of DF and CF for fixed-power background noises, while [41] provides analytical expressions for the end-to-end SNR and OP of cooperative diversity in correlated lognormal channels for full-duplex DF relaying, and selection combining or maximum ratio combining, as a function of the joint channel densities.

There has been much work in the context of wireless networks with the aid of stochastic geometry models. The simplest of these models, the one which has been most frequently used is the homogeneous PPP. The fundamental benefit of this process is that it generally leads to mathematically simpler and more tractable expressions than other, more structured processes. Initially, the PPP model for node distribution was taken to be a reasonable approximation for decentralized networks in which transmissions take place in uncoordinated fashion, such as with the ALOHA medium access strategy. Examples of these are wireless sensor and *ad hoc* networks [10, 42]. In recent years, it has also been adopted for more structured networks such as cellular networks, where macro, pico, femto base station distributions are modeled via this process [43, 44]. It is shown that, even though this assumption implies the independent distribution of BSs, which may not hold in cellular networks, the model is still accurate and conservative in predicting many important network parameters, while retaining a higher degree of mathematical tractability than standard regular-grid models.

In the context of cellular networks, cooperative schemes with spatial models have mainly regarded cooperation between two or more base stations with a wired backhaul (up to the moment in which the results in this and the following chapter were developed). This has been motivated mainly by the new cooperation modes introduced in recent LTE releases. In [45], the authors analyze the improvements in coverage probability, achievable through a scheme in which each user can be served by the base stations which is nearest to him, or jointly served by the two nearest ones. On the other hand, in [46], the authors use a model based on the Poisson process to analyze the interference distribution in a network employing non-coherent joint-transmission base station cooperation, in which several base stations transmit the same data to a given user, without prior phase mismatch correction. They rely on approximating the co-channel interference experienced by a reference receiver by a Gamma random variable with the same moments as the true interference, an approach which was proposed in [47]. Finally, in [48], the authors perform an analysis of base station cooperation (through CoMP) in cellular networks modeled as PPPs without approximating the interference.

5.1.2 Contributions

We focus on a reference transmitter which attempts to communicate with a destination with the aid of a full-duplex in-band relay, using either the DF or CF schemes. This channel could be interpreted as the downlink of a reference fixed-size cell of a cellular network, in which the BS cooperates with a wireless infrastructure relay. The relay is therefore not connected to the wired backhaul of the network, as in previous works, and receives the message wirelessly from the BS. We consider that the co-channel interference experienced by the relay and the destination of the message comes from nodes which are distributed as a PPP and that signals are subject to path loss attenuation and Rayleigh fading. In this analysis, we focus on the interference and the SIR, neglecting the presence of background noise, since in many scenarios this is the main limit to performance [43]. For DF, we derive the expression of the exact OP of the link in terms of the joint Laplace transform of the interferences, which unfortunately, can only be evaluated numerically. In addition, we perform an asymptotic analysis of the OP as the density of interferers tends to zero (SIR $\to \infty$), and show that it is minimized when the correlation coefficient between the source and relay's symbols is zero. We then derive closed-form upper bounds to the OP which are tight for small OPs, typical in wireless system designs [24]. In the case of DF, the requirement that the relay fully decode the messages of the source prior to forwarding them, implies that the rate that the link can achieve does not depend on the spatial correlation of the interference signals at the relay and the destination. This spatial correlation is a consequence of the spatial distribution of interfering nodes which interfere with the relay and the destination simultaneously. In CF, however, the relay compresses the messages received from the source without decoding them and hence, the achievable rate depends on the spatial correlation induced by the point process. For this reason, a direct analysis of the CF protocol is infeasible. Here, we show that the achievable rate of CF with correlated interference is at most one bit worse than with uncorrelated interference and, using this fact, we derive an upper bound on the OP of CF. Finally, we also compare the performance of the analyzed protocols with a simple half-duplex DF protocol and with DT. To do this, a lower bound on the OP for a half-duplex DF protocol is introduced.

The chapter is organized as follows: in Section 5.2, we present the mathematical model of the network, a description of the DF and CF schemes, and their achievable rates. The OP is also shown to be an upper bound on the asymptotic error probability of the link. In Section 5.3 the OP analysis is carried out for both schemes. Finally, numerical results and conclusions can be found in Sections 5.4 and 5.5, respectively, while proofs are relegated to Appendix B.

Notation

In this chapter we use the big O notation: $f(x) = O\left(g(x)\right)$ as $x \to x_0$ if there exists $M > 0$ such that $|f(x)| \leq M|g(x)|$ in some neighborhood of x_0.

5.2 General considerations and network model

5.2.1 Spatial model of node distribution

We consider a single-antenna source node located at the origin $o \in \mathbb{R}^2$ which attempts to communicate to a destination located at $d = (D, 0)$, with the aid of a relaying node located at r and working in a full-duplex fashion in the same frequency band. We assume a narrow-band slow-fading scenario[1] and that there is no node mobility; this implies that during the transmission time all the positions of the nodes and other network parameters, including fading coefficients, remain constant. With this, we model the interfering nodes as an independently marked homogeneous PPP (Sec. 4.3):

$$\tilde{\Phi} = \left\{(x_i, h_{x_i r}, h_{x_i d})\right\}, \tag{5.1}$$

with the following characteristics:

- The set of transmitters constitutes an homogeneous PPP $\Phi = \{x_i\}$ of intensity λ.
- All users transmit with constant unit power. We assume that transmissions are affected by path loss and i.i.d. narrow-band block-fading, that is, the power received at y by a transmitter at x is $|h_{xy}|^2 l(x, y)$ where:
  - $l(x, y)$ is a spherically symmetric path loss between x and y. For numerical results we shall work with the usual unbounded or *simplified* path loss function: $l(x, y) = ||x - y||^{-\alpha}$ with $\alpha > 2$.
  - $|h_{xy}|^2$ is the power fading coefficient associated with the channel between points x and y. We consider Rayleigh fading, i.e. the power fading coefficients are independent identically distributed exponential random variables with unit mean. This is equivalent to saying that h_{xy} are zero-mean CCSG random variables [6].
- The marks h_{xr} and h_{xd} model the fading coefficient between each transmitting node in the network and the nodes relay and destination corresponding to the transmitter located at the origin, respectively. In addition we include another fading coefficient h_{rd}

[1] For an summary on the motivation and characteristics of slow-fading frequency non-selective channels see Appendix A.

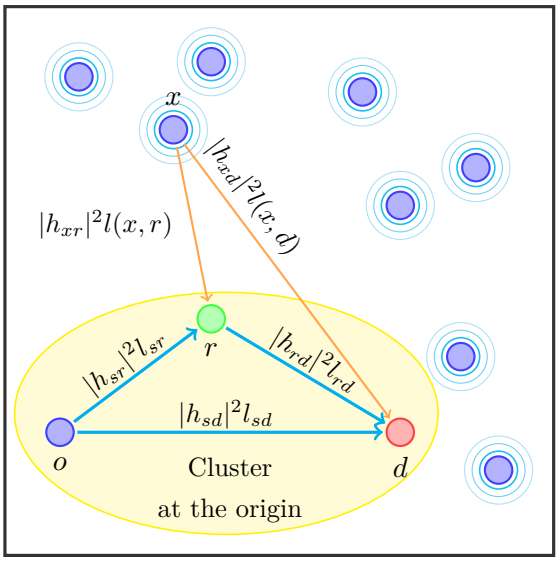

Figure 5.1: Representation of the relevant power fading coefficients of the relay channel at the origin and its interfering nodes (in blue).

with the same distribution as h_{xr} and h_{xd}, independent of $\tilde{\Phi}$, which models the fading between the relay and destination corresponding to the transmitter at the origin. We denote by l_{sd}, l_{sr} and l_{rd} the source-destination, source-relay and relay-destination path losses respectively.

In Fig. 5.1 we can see a representation of all the relevant variables defined. The signals received at the relay and destination, associated with the source node at the origin, can be written as:

$$Y_r = h_{sr}\sqrt{l_{sr}}X_s + \underbrace{\sum_{x\in\Phi} h_{xr} l(x,r)^{\frac{1}{2}} X_x}_{\tilde{Z}_r},$$

$$Y_d = h_{sd}\sqrt{l_{sd}}X_s + h_{rd}\sqrt{l_{rd}}X_r + \underbrace{\sum_{x\in\Phi} h_{xd} l(x,d)^{\frac{1}{2}} X_x}_{\tilde{Z}_d},$$

where, for shortness, we have dropped the dependence of the signals on the messages to be transmitted and the discrete time indices for the block codewords. We have denoted with (X_s, X_r) the symbols transmitted by the source and the relay, and $\{X_{x_i}\}_i$ the corresponding signals for the other transmitters in the network. If the interfering users employ Gaussian signaling, that is, the $\{X_{x_i}\}_i$ are generated as unit-variance zero-mean independent CCSG random variables, then for each realization of $\tilde{\Phi}$ the aggregate interferences $\tilde{Z}_r$ and $\tilde{Z}_d$, are

zero-mean CCSG variables whose conditional variances are:

$$I_r \triangleq \mathbb{E}\left[|\tilde{Z}_r|^2|\tilde{\Phi}\right] = \sum_{x\in\Phi} |h_{xr}|^2 l(x,r), \tag{5.2}$$

$$I_d \triangleq \mathbb{E}\left[|\tilde{Z}_d|^2|\tilde{\Phi}\right] = \sum_{x\in\Phi} |h_{xd}|^2 l(x,d). \tag{5.3}$$

In addition, these time signals are spatially correlated, and their correlation coefficient is:

$$\rho_N \triangleq \frac{\mathbb{E}[\tilde{Z}_r \tilde{Z}_d^*|\tilde{\Phi}]}{\sqrt{I_r I_d}}. \tag{5.4}$$

This correlation coefficient appears because the interference time signals are generated at r and d by the same set of transmitters, and is not due to the randomness of the point process (it would exist even if the nodes where fixed and the point process wasn't considered). This means that, for every realization of $\tilde{\Phi}$ and each time instant, if the interference is treated as noise for decoding, the interference signals at r and d can be treated as correlated Gaussian noise whose correlation coefficient is given by (5.4). As we mentioned previously, depending on the protocol used, this correlation may have an impact on the achievable rate and it may be necessary to take it into account in the analysis.

The randomness of the distribution of nodes, modeled by the point processes, implies that the interference powers and the interference correlation (5.4) will be random variables, and will themselves be correlated. We can define the joint Laplace transform for these interference power random variables I_r and I_d as:

$$\mathcal{L}_{I_d,I_r}(\omega_1,\omega_2) \triangleq \mathbb{E}_{\tilde{\Phi}}\left[e^{-(\omega_1 I_d + \omega_2 I_r)}\right],\ \omega_1,\omega_2 \in \mathbb{C}, \tag{5.5}$$

with $\Re\{\omega_1\}, \Re\{\omega_1\} > 0$. In this case, given the expressions of the interferences (5.2) and (5.3) this transform is also the Laplace functional (Def. 4.2.9) of the point process with respect to the function:

$$\tilde{f}(x,h_1,h_2,r,d,\omega_1,\omega_2) = \omega_1 g(x,d,h_1) + \omega_2 g(x,r,h_2), \tag{5.6}$$

where $g(x,y,h) = l(x,y)|h|^2$ is the function appearing inside the interferences. Therefore, using (4.14), we get:

$$\mathcal{L}_{I_d,I_r}(\omega_1,\omega_2) = \exp\left\{-\lambda \int_{\mathbb{R}^2} \mathbb{E}_{h_1,h_2}\left[1 - e^{-\omega_1 g(x,d,h_1) - \omega_2 g(x,r,h_2)}\right] dx\right\}, \tag{5.7}$$

$$= \exp\left\{-\lambda \int_{\mathbb{R}^2} \left[1 - \frac{1}{(1+\omega_1 l(x,d))(1+\omega_2 l(x,r))}\right] dx\right\}. \tag{5.8}$$

In the second step we computed the expectation over $\{|h_1|^2, |h_2|^2\}$, which are unit mean independent exponential random variables representing the fading coefficients inside the interference random variables.

Lemma 5.1. *For the simplified path loss function the Laplace transform can be written as:*

$$\mathcal{L}_{I_d,I_r}(\omega_1,\omega_2) = e^{-\lambda\left(\tilde{C}(\omega_1^{2/\alpha}+\omega_2^{2/\alpha})+f(\omega_1,\omega_2)\right)}, \tag{5.9}$$

where:

$$\tilde{C} \triangleq \frac{2\pi\Gamma\left(\frac{2}{\alpha}\right)\Gamma\left(1-\frac{2}{\alpha}\right)}{\alpha}, \tag{5.10}$$

$$f(\omega_1,\omega_2) = \int_{\mathbb{R}^2} \frac{\omega_1\omega_2}{(\omega_1+\|x-d\|^\alpha)(\omega_2+||x-r||^\alpha)}dx, \tag{5.11}$$

and $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$ *is the Gamma function.*

Proof. See Appendix B.2. □

$f(\omega_1,\omega_2)$ in (5.11) accounts for the statistical dependence (induced by the point process) between the interference powers at two different locations. If, in fact, they were independent, the joint Laplace transform would be the product of their individual transforms and this cross-term would not appear. Unfortunately this term does not have a closed form and is very difficult to bound tightly in a general setup for all (ω_1,ω_2). This is because the correlation in the interference powers will change a lot if the relay and the destination are close or far away, meaning that the function will change a lot, which makes finding a general bound or approximation very hard. Taking $\omega_1 = 0$ or $\omega_2 = 0$, the separate Laplace transforms of the interference powers, which have closed form expressions for the simplified path loss function, can be obtained.

5.2.2 Problem statement and bounds on the asymptotic error probability

Our goal is to study the asymptotic error probability performance of the relay channel formed by the source at the origin together with its destination and the relay. In order to study this,

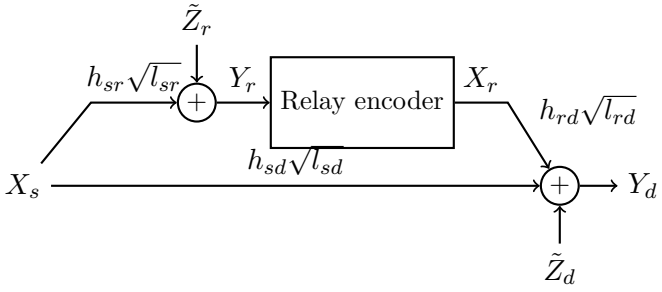

Figure 5.2: Block diagram of the relay channel.

we introduce a family of coding schemes which the three users composing the relay channel

are allowed to use. In this way a notion of optimality in terms of error probability, within the set of admissible coding schemes, can be defined.

Definition 5.2.1 (Single-relay code)**.** *A single-relay code $\mathcal{C}_n(n, 2^{nR})$ for a set of messages $\{1, \ldots, 2^{nR}\}$ consists of:*

- *A set of codewords $X_s^n(w_i) \in \mathbb{C}^n$, $w_i \in \{1, \ldots, 2^{nR}\}$, drawn independently, each according to n independent draws of a unit-variance CCSG random variable.*
- *A decoder mapping $\hat{W} : \mathbb{C}^n \longmapsto \{1, \ldots, 2^{nR}\} \cup \{\mathcal{E}\}$.*
- *A sequence of relay mappings $f_t : \mathbb{C}^{t-1} \longmapsto \mathbb{C}$ constrained to produce independent complex Gaussian random variables of unit variance, for $t = \{1, \ldots, n\}$.*

A representation of the relay channel is shown in Fig. 5.2. In order to perform a transmission, a single-relay code is selected and given to the three users. A message $W \in \{1, \ldots, 2^{nR}\}$ is chosen at random by the source, which then transmits its corresponding codeword $X_s^n(W)$. At time instant i, the relay outputs $X_{r,i} = f_i(Y_r^{i-1})$, where $Y_r^{i-1} = (Y_{r,1}, \ldots, Y_{r,i-1})$ are its previously received symbols. Finally, the destination uses the decoder mapping to estimate the sent message W using $Y_d^n = (Y_{d,1}, \ldots, Y_{d,n})$.

For a given rate R, the smallest asymptotic average probability of error that can be obtained through a sequence of codes $\{\mathcal{C}_n\}_n$ as in Def. 5.2.1 is given by:

$$\bar{P}_e(R, \lambda) \triangleq \inf_{\{\mathcal{C}_n\}_n} \Big\{ \limsup_{n\to\infty} \mathbb{P}_\Theta^{(n)}(W \neq \hat{W}|\mathcal{C}_n) \,\Big|\, \liminf_{n\to\infty} \frac{1}{n} \log M_n \geq R \Big\}, \tag{5.12}$$

where Θ condenses all the randomness in the model:

$$\Theta = \left\{ \tilde{\Phi}, h_{sr}, h_{rd}, h_{sd} \right\}. \tag{5.13}$$

The error probability depends on several additional parameters, for example, the path loss exponent, but we prefer to make explicit its dependence on the rate R. We assume there is no channel state information (CSI) available at the source, that the relay only has CSI of the source-relay channel, and the destination of both the source-destination and relay-destination channels. Since the source is unaware of the instantaneous interference, path loss attenuation, and fading coefficients involved, it cannot adapt its transmission rate, and hence the error probability cannot be made arbitrary small with the code-length. For any code $\mathcal{C}_n$ according to Def. 5.2.1, the error probability can be bounded as [24]:

$$\begin{aligned} \mathbb{P}(W \neq \hat{W}|\mathcal{C}_n) &= \mathbb{P}(W \neq \hat{W}|\mathcal{C}_n, \mathcal{O}(R))\ \mathbb{P}(\mathcal{O}(R)|\mathcal{C}_n) + \mathbb{P}(W \neq \hat{W}|\mathcal{C}_n, \mathcal{O}^c(R))\ \mathbb{P}(\mathcal{O}^c(R)|\mathcal{C}_n) \\ &\leq \mathbb{P}(\mathcal{O}(R)|\mathcal{C}_n) + \mathbb{P}(W \neq \hat{W}|\mathcal{C}_n, \mathcal{O}^c(R)), \end{aligned} \tag{5.14}$$

where $\mathcal{O}(R) \in \sigma(\Theta)$ denotes an outage event, $\sigma(\Theta)$ being the σ-algebra generated by Θ. Therefore, the asymptotic error probability can be upper bounded by any sequence of codes $\{\mathcal{C}_n\}_n$, as follows:

$$\bar{P}_e(R,\lambda) \leq \inf_{\mathcal{O}(R)\in\sigma(\Theta)} \left[\mathbb{P}_\Theta\left(\mathcal{O}(R)\right) + \limsup_{n\to\infty} \mathbb{P}_\Theta^{(n)}\left(W \neq \hat{W} | \mathcal{C}_n, \mathcal{O}^c(R)\right) \right], \tag{5.15}$$

where $\mathbb{P}_\Theta\left(\mathcal{O}(R)\right)$ is the asymptotic probability as $n \to \infty$ of the outage event. If, for a given sequence $\{\mathcal{C}_n\}_n$, the event $\mathcal{O}^c(R)$ is chosen to contain the points in Θ such that R is achievable through $\{\mathcal{C}_n\}_n$, the second term on the right-hand side of (5.15) can be made arbitrary small. That is, for any $\epsilon > 0$:

$$\limsup_{n\to\infty} \mathbb{P}_\Theta^{(n)}\left(W \neq \hat{W} | \mathcal{C}_n, \mathcal{O}^c(R)\right) \leq \epsilon. \tag{5.16}$$

In this way, given a rate R, the asymptotic error probability $\bar{P}_e(R,\lambda)$ is dominated by the OP $\mathbb{P}_\Theta\left(\mathcal{O}(R)\right)$ of the corresponding achievable rate. The OP is a useful performance metric which was extensively employed to characterize the performance of networks in a Poisson field of interferers, jointly with other associated metrics like the *transmission capacity* (TC) [30, 42]. The transmission capacity was introduced in [49] in order to include outage probability constraints in the scaling behavior. Several results have been obtained, through the use of the TC, for several practical situations, as multiple input-multiple output capable users in wireless networks [50], decentralized power control [51], etc.

In what follows, for shortness we write $\mathbb{P}_\Theta \equiv \mathbb{P}$. We shall also consider the scaling behavior of the error probability with the density of interferers. We have the following definition:

Definition 5.2.2 (Small node-density regime)**.** *The following metric $\kappa(R)$ characterizes the asymptotic error probability $\bar{P}_e(R,\lambda)$ as $\lambda \to 0$, that is, in the high-SIR [52, 53], or small node-density regime:*

$$\kappa(R) \triangleq \lim_{\lambda\to 0} \frac{\bar{P}_e(R,\lambda)}{\lambda}. \tag{5.17}$$

This parameter indicates the behavior of the error probability as the density of interferers tends to zero (SIR $\to \infty$), in which case we have, $\bar{P}_e(R,\lambda) \approx \kappa(R)\lambda$. This is a good approximation in the typical, small error probability operating regime of wireless networks [24]. Since the best error probability, and hence, $\kappa(R)$, cannot be found in closed form, we can bound it as:

$$\kappa(R) \leq \lim_{\lambda\to 0} \frac{\mathbb{P}\left(\mathcal{O}(R,\lambda)\right)}{\lambda}. \tag{5.18}$$

The quantity on the right side of (5.18), introduced in [52], is known as the *spatial contention* parameter, and it represents the slope of the OP as $\lambda \to 0$. It is therefore an upper bound to the best slope attainable, given by $\kappa(R)$. In fact, in [52] the authors study the outage

behavior of general motion-invariant networks employing DT, by resorting to an asymptotic analysis in which the density of interfering nodes goes to zero. In particular they show that the OP using an arbitrary medium access scheme approaches $\gamma\lambda^{\varsigma}$ as the density of interferers $\lambda \to 0$, where γ is the *spatial contention* parameter and ς is the *interference scaling exponent*. For the case of networks using the ALOHA access scheme or in this case, we have that $\varsigma = 1$. The *high reliability regime* as defined in [52] refers to the operating regime in which the OP is small enough (close to zero) to guarantee that the asymptotic first order approximation is a good representation of the network performance. This regime covers OP values of the order of 0.01 which are typical in wireless system designs [3].

5.2.3 Achievable bounds on the asymptotic error probability

As mentioned before, Cover and El Gamal introduced the main coding strategies for the relay channel in their seminal paper [14]: DF and CF. The general capacity for the relay channel is not known, so in general there is not a unique scheme maximizing the rate for all channel parameters.

Decode-and-Forward

In this protocol, the relay decodes the messages sent by the source, re-encodes them and forwards them to the destination, which employs the transmissions of both users to decode the message. In the special case where the memoryless relay channel is *physically degraded* the achievable rate using DF is in fact the capacity. Since the standard version of DF requires the relay to fully decode the message of the source, this strategy will work best when this channel is good enough with respect to the source-destination channel so that a bottleneck is not introduced. In a scenario in which the spatial attenuation of signals is considered, this will happen when the relay is on average closer to the source than to the destination. Other variants of DF such as partial decode-and-forward [54] partially overcome this requirement, but they require an optimization of the code at the encoder, which cannot be done unless CSI is available at the source.

In order to bound the average asymptotic error probability with $\mathbb{P}(\mathcal{O}(R))$ as discussed in the previous section, we need to define the outage events associated with DF. There exist several coding schemes which achieve the same DF rate, all based on block-Markov coding [33]. Even though all the schemes achieve the same rate, the different encoding/decoding procedures result in different outage events, and hence, different outage probabilities. Here we consider the outage events associated to DF with *regular encoding* and *sliding-window decoding* [55] at the destination, or *regular encoding* and *backward decoding* [56, 57]. In

Block	1	2	3	...	B-1	B
S. transmits	$x_s^n(w_1\|1)$	$x_s^n(w_2\|w_1)$	$x_s^n(w_3\|w_2)$	...	$x_s^n(w_{B-1}\|w_{B-2})$	$x_s^n(1\|w_{B-1})$
R decodes	$\tilde{w}_1$	$\tilde{w}_2$	$\tilde{w}_3$	...	$\tilde{w}_{B-1}$	$\emptyset$
R. transmits	$x_r^n(1)$	$x_r^n(\tilde{w}_1)$	$x_r^n(\tilde{w}_2)$	...	$x_r^n(\tilde{w}_{B-2})$	$x_r^n(\tilde{w}_{B-1})$
D. decodes	$\emptyset$	$\hat{w}_1$	$\hat{w}_2$	...	$\hat{w}_{B-2}$	$\hat{w}_{B-1}$

Figure 5.3: Encoding procedure for DF with regular encoding and sliding window decoding.

[14], DF is defined using *irregular encoding*, *random binning* and *successive decoding* at the destination but this strategy has additional error events so it will not be considered[2]. Table 5.3 shows the encoding and decoding structure of the protocol under sliding window decoding. The transmission is split into B blocks, in which independent messages are transmitted, with $w_0 = w_B = 1$. At the end of the block b, the relay decodes the message sent by the source in the same block, which it then transmits in block $b+1$. At the end of block $b+1$, the destination decodes the message sent by the source during block b, using also the transmission of the relay in block $b+1$. Using Gaussian signaling, the n-length random codewords at each source and its associated relay are:

$$X_s^n(w_{i-1}, w_i) = \sqrt{(1-|\rho|^2)}\tilde{X}_1^n(w_i) + \rho\tilde{X}_2^n(w_{i-1}), \tag{5.19}$$

$$X_r^n(w_{i-1}) = \tilde{X}_2^n(w_{i-1}), \tag{5.20}$$

for messages $w_i \in \{1, \ldots, 2^{nR}\}$ with $w_0 = w_B = 1$ and each block $i = \{1, \ldots, B\}$. $\tilde{X}_1$ and $\tilde{X}_2$ are independent zero-mean CCSG random variables with unit variance and ρ is the correlation coefficient between source and relay signals X_s and X_r. This DF protocol also has the property of being oblivious to the presence of the relay, that is, if the relay does not decode or choses not to transmit, the destination can still use the transmission from the source to decode the message, so the performance of the protocol is not degraded with respect to DT [38, 58]. The destination needs to know if the relay will transmit, which is cost-free as the block-length grows. The relay will be able to decode the source's transmission if the attempted rate R satisfies:

$$R < I(X_s; Y_r|X_r), \tag{5.21}$$

where $I(\cdot;\cdot|\cdot)$ denotes conditional mutual information [4]. Since the nodes use Gaussian

[2]In this version of DF, the relay groups the messages into bins. After receiving the transmission from the source, it retransmits the bin index corresponding to the received message instead of the message itself. The destination then decodes the bin index and uses the index to decode the message using the transmission from the source. In an outage framework, we have the outage events of the two decoding steps at the destination. In the other DF versions, the decoder jointly decodes using the transmission from the source and the relay in one step, so there is only one outage event associated to the destination.

signaling, that is, the codewords are generated according to (5.19) and (5.20), evaluating the mutual information (5.21) (see Appendix B.1) the event that the relay is unable to decode source's message is:

$$\mathcal{A}_{DF}(R,\rho) \triangleq \left\{ \mathsf{C}\left(\frac{|h_{sr}|^2 l_{sr}(1-|\rho|^2)}{I_r} \right) < R \right\}, \tag{5.22}$$

with $\mathsf{C}(u) = \log_2(1+u)$. When the relay decodes the message, the achievable rate is that of a two-antenna transmission towards the destination, which achieves any rate $R < I(X_r, X_s; Y_d)$. On the other hand, since the protocol is oblivious, in the case that the relay does not decode (event (5.22)) takes place, any rate $R < I(X_s; Y_d)$ is achievable. Then for each realization of $\tilde{\Phi}$ any attempted rate R that satisfies:

$$R \leq \mathbb{1}\{\mathcal{A}_{DF}(R,\rho)\} R_M(\rho) + \mathbb{1}\{\mathcal{A}^c_{DF}(R,\rho)\} R_{DT}, \tag{5.23}$$

is achievable, where:

$$R_{DT} = \mathsf{C}\left(\frac{|h_{sd}|^2 l_{sd}}{I_d} \right), \tag{5.24}$$

$$R_M(\rho) = \mathsf{C}\left(\frac{|h_{sd}|^2 l_{sd} + |h_{rd}|^2 l_{rd} + 2\sqrt{l_{sd} l_{rd}} \Re\left(\rho h_{sd} h^*_{rd}\right)}{I_d} \right), \tag{5.25}$$

are the rates of a DT from the source to the destination, and a joint transmission from the source and the relay to the destination, respectively. When $\rho = 0$, the rates of this protocol are the same as those achievable through a scheme known as block-Markov multiplexed coding [59, 60]. Introducing the following outage events:

$$\mathcal{B}_{DF}(R,\rho) \triangleq \{R_M(\rho) < R\}, \tag{5.26}$$

$$\mathcal{A}_{DT}(R) \triangleq \{R_{DT} < R\}, \tag{5.27}$$

the outage event for this protocol is, from (5.23):

$$\mathcal{O}_{DF}(R,\rho) \triangleq [\mathcal{A}^c_{DF}(R,\rho) \cap \mathcal{B}_{DF}(R,\rho)] \cup [\mathcal{A}_{DF}(R,\rho) \cap \mathcal{A}_{DT}(R)], \tag{5.28}$$

for which condition (5.16) holds true. The event $\mathcal{B}_{DF}(R,\rho)$ means that the destination is in outage while source and relay cooperate. The error probability is bounded by:

$$\bar{P}_e(R,\lambda) \leq \inf_{\rho \in \mathbb{C}, |\rho| \leq 1} P_{\text{out,DF}}(R,\rho), \tag{5.29}$$

where $P_{\text{out,DF}}(R,\rho) = \mathbb{P}\left(\mathcal{O}_{DF}(R,\rho)\right)$. Notice that the imposition of full decoding at the relay implies that the achievable rate does not depend on the correlation of the interference time signals at the relay and destination, given by (5.4).

Block	1	2	3	...	$B-1$	B
S transmits	$x_s^n(w,1)$	$x_s^n(w,2)$	$x_s^n(w,2)$	...	$x_s^n(w,B-1)$	$x_s^n(w,B)$
R decodes	$\hat{y}_r^n(l_1\|1),l_1$	$\hat{y}_r^n(l_2\|l_2),l_2$	$\hat{y}_r^n(l_3\|l_2),l_3$	...	$\hat{y}_r^n(l_{B-1}\|l_{B-2}),l_{B-1}$	$\emptyset$
R transmits	$x_r^n(1)$	$x_r^n(\tilde{l}_1)$	$x_r^n(\tilde{l}_2)$	...	$x_r^n(\tilde{l}_{B-2})$	$x_r^n(\tilde{l}_{B-1})$
D decodes	$\emptyset$	$\emptyset$	$\emptyset$	...	$\emptyset$	$\hat{w}$

Figure 5.4: Compress and forward (noisy network coding scheme) transmission and decoding strategy.

Compress-and-Forward

In this scheme, the relay compresses the received signal without decoding the message and forwards this compressed description to the destination. In the same was as with DF, there are several schemes which achieve the same CF rate and have different outage events and different outage probabilities. In this work we shall consider the outage events induced by noisy network coding [61, Sec. III]. In Table 5.4 we can see a representation of the encoding and decoding procedure. The whole transmission is split into B blocks, in which the source transmits the same message B times using independent codebooks. At the end of the block b, the relay finds a reconstruction sequence $\hat{y}_2^n(b)$ of its received sequence $y_2^n(b)$, conditioned on its transmission during the block, $x_2^n(b)$. Instead of transmitting the sequence $\hat{y}_2^n(b)$, the relay will use a binning technique and transmit a bin-index l_b corresponding to $\hat{y}_2^n(b)$ in the following block. The destination waits until the end of the B-th block and determines the transmitted message which is compatible with all the transmissions from the source and the relay during the B blocks. This protocol has two outage events: that the relay does not decode the right index at the end of each block, and that the destination fails to decode the message at the end of the B blocks. Considering these two events and chosen a probability function $p_{X_s}p_{X_r}p_{\hat{Y}_r|X_r,Y_r}$ the rate than can be achieved is [61]:

$$R_{CF} = \min\{I(X_s,X_r;Y_d) - I(Y_r;\hat{Y}_r|X_s,X_r,Y_d), I(X_s;\hat{Y}_r,Y_d|X_r)\}. \tag{5.30}$$

$\hat{Y}_r$ represents the compressed representation of the symbols received by the relay Y_r. As the relay is not compelled to decode the source message, there is not a bottleneck in the information flow through the relay as in DF. When the relay is close to the destination, CF will compress Y_r and transmit this description to the destination with little effort and CF will typically outperform DF and DT. In the Gaussian relay channel it is usual [33] to choose $\hat{Y}_r = Y_r + Z_c$ and X_s, X_r and Z_c independent zero-mean CCSG random variables with unit variance for the first two and variance n_c for the third one. Thus, the following rate can be

achieved from (5.30) (see Appendix B.1 for an explanation on how these rates are derived):

$$R_{CF}(\rho_N, n_c) = \min\{R_1(\rho_N, n_c), R_2(\rho_N, n_c)\}, \tag{5.31}$$

$$R_1(\rho_N, n_c) = \mathsf{C}\left(\frac{|h_{sd}|^2 l_{sd} + |h_{rd}|^2 l_{rd}}{I_d}\right) - \mathsf{C}\left(\frac{I_r}{n_c}(1 - |\rho_N|^2)\right)$$

$$R_2(\rho_N, n_c) = \mathsf{C}\left(\frac{|h_{sd}|^2 l_{sd}}{I_d} + \frac{\frac{|h_{sd}|^2 l_{sd}}{I_d}|\rho_N|^2 + \frac{|h_{sr}|^2 l_{sr}}{I_r} - 2\Re\left\{\rho_N \frac{h_{sd}\sqrt{l_{sd}}}{\sqrt{I_r}} \frac{h_{sr}^*\sqrt{l_{sr}}}{\sqrt{I_d}}\right\}}{1 + \frac{n_c}{I_r} - |\rho_N|^2}\right). \tag{5.32}$$

We have made explicit the dependence of the achievable rate with the spatial noise correlation coefficient ρ_N given by (5.4) to mark a distinction with DF in which this correlation does not affect the rate. Notice that the rate is also dependent on the compression variance n_c of choice. In general, whenever:

$$I(X_r; Y_d) \geq I(Y_r; \hat{Y}_r | X_r, Y_d), \tag{5.33}$$

then the rate R_{CF} is the second term in (5.31) [54], that is:

$$R_{CF}(\rho_N) = I(X_s; \hat{Y}_r, Y_d | X_r). \tag{5.34}$$

Instead of considering (5.31) we can use (5.33) and (5.34) which are simpler to work with (though they give a smaller rate)[3]. For the Gaussian relay channel, after choosing X_s, X_r and Z_c as indicated above (5.31), condition (5.33) is [14]:

$$n_c \geq \frac{I_r I_d}{|h_{rd}|^2 l_{rd}}\left(\frac{|h_{sd}|^2 l_{sd}}{I_d} + \frac{|h_{sr}|^2 l_{sr}}{I_r} - 2\Re\left\{\rho_N \frac{h_{sd} h_{sr}^* \sqrt{l_{sd} l_{sr}}}{\sqrt{I_r I_d}}\right\} + 1\right). \tag{5.35}$$

This means that is we define the event $\mathcal{B}_{CF}(n_c, \rho_N) = \{(5.35) \text{ is not met}\}$, then any rate R that satisfies:

$$R < \mathbb{1}\{\mathcal{B}_{CF}^c(n_c, \rho_N, n_c)\} R_2(\rho_N, n_c), \tag{5.36}$$

is achievable [14]. Condition (5.35) implies that the relay-destination channel can sustain the rate to transmit the compressed version of what the relay receives from the source. If full CSI is available at the relay, we can choose the compression variance n_c to achieve equality in (5.35). Otherwise, the value of n_c has to be fixed a priori and an outage event will take place when the realization of the network does not allow (5.35) to be fulfilled. Therefore, we can define the outage event $\mathcal{O}_{CF}(R, \rho_N, n_c) = \{\mathcal{A}_{CF}(R, \rho_N, n_c) \cup \mathcal{B}_{CF}(n_c, \rho_N)\}$, with:

$$\mathcal{A}_{CF}(R, \rho_N, n_c) = \{R_2(\rho_N, n_c) < R\}. \tag{5.37}$$

Notice that CF cannot perform worse than DT, because by taking an arbitrarily large value of n_c we guarantee that (5.35) will be met, and in that case, inspecting (5.32) we check that $R_2(\rho_N, n_c)$ will be very arbitrarily close to the rate of DT.

[3] Actually, in traditional scenario the rate regions defined by (5.31), and by (5.33) and (5.34) are maximized over all distributions $p_{X_s} p_{X_r} p_{\hat{Y}_r|X_r,Y_r}$ and in that case the regions rates are the same [23, 54].

Half-duplex DF

In order to compare the performance of the full-duplex DF protocol we introduced before, we also consider a half-duplex DF strategy, known as *sequential* DF [38]. In this scheme, the transmission is split in two phases. In the first one, occupying a fraction $0 \leq \varepsilon < 1$ of the block, the source transmits its message to the destination while the relay listens and attempts to decode the message. If the relay is able to decode the message during the first phase, it employs the remaining $(1-\varepsilon)$ fraction of the block to transmit, acting as a secondary antenna. This scheme is also oblivious, so if the relay does not decode the message, it does not degrade the performance with respect to DT. The event that the relay does not decode in the first phase is:

$$\mathcal{A}_{SDF}(R,\varepsilon) \triangleq \left\{ \varepsilon \ \mathsf{C}\left(\frac{|h_{sr}|^2 l_{sr}}{I_r}\right) < R \right\}. \tag{5.38}$$

Hence, any rate R which satisfies:

$$R < \mathbb{1}\{\mathcal{A}^c_{SDF}(R,\varepsilon)\} R_{SDF}(\varepsilon) + \mathbb{1}\{\mathcal{A}_{SDF}(R,\varepsilon)\} R_{DT},$$

is achievable, where R_{DT} is given by (5.24) and [38]:

$$R_{SDF}(\varepsilon) = \varepsilon \ \mathsf{C}\left(\frac{|h_{sd}|^2 l_{sd}}{I_d}\right) + (1-\varepsilon)\mathsf{C}\left(\frac{|h_{sd}|^2 l_{sd} + |h_{rd}|^2 l_{rd}}{I_d}\right).$$

The outage event for this protocol is therefore:

$$\mathcal{O}_{SDF}(R,\varepsilon) \triangleq [\mathcal{A}^c_{SDF}(R,\varepsilon) \cap \mathcal{B}_{SDF}(R,\varepsilon)] \cup [\mathcal{A}_{SDF}(R,\varepsilon) \cap \mathcal{A}_{DT}(R)], \tag{5.39}$$

where $\mathcal{B}_{SDF}(R,\varepsilon) = \{R_{SDF}(\varepsilon) < R\}$. The value of ε cannot be adjusted for the instantaneous realization of the network, but can be selected a priori, for example, to minimize the OP.

Direct transmission

We also define the outage event $\mathcal{O}_{DT}(R) \triangleq \mathcal{A}_{DT}(R)$ given by (5.27) for the case in which there is no relay and thus the source simply uses DT. The OP for this scheme is known to be [10]:

$$P_{\text{out,DT}} = \mathbb{P}(\mathcal{O}_{DT}(R)) = 1 - \exp\{-\lambda \delta D^2\}, \tag{5.40}$$

where, using $\tilde{C}$ given by (5.10), we defined:

$$\delta \triangleq \tilde{C} T(R)^{2/\alpha}, \tag{5.41}$$

$$T(R) \triangleq 2^R - 1. \tag{5.42}$$

5.3 Outage behavior

In this section we analyze the outage behavior of the relay channel. In the case in which only Gaussian background noise and Rayleigh fading are considered (without interference) very interesting gains have been observed in terms of the OP [15, 16]. In the scenario in which interference comes from a network of interferers, however, we must average over all possible configurations of interfering nodes, considering numerous situations in which communications are severely impaired due to the presence of heavy interference so the performance gains may not be so large.

5.3.1 Decode-and-forward

We start by considering the DF protocol. In this setup we derive the OP and tight upper bounds under typical network operating conditions. We also determine the symbol correlation coefficient ρ which minimizes the OP in the small node-density regime. To analyze the OP (5.29), it is convenient to rewrite in terms of the success events as:

$$P_{\text{out,DF}}(R,\rho) = 1 - \mathbb{P}\left(\mathcal{A}_{DF}^{c}(R,\rho) \cap \mathcal{B}_{DF}^{c}(R,\rho)\right) - \mathbb{P}\left(\mathcal{A}_{DF}(R,\rho) \cap \mathcal{A}_{DT}^{c}\right). \tag{5.43}$$

It is interesting to mention that the probability of $\mathcal{B}_{DF}$ in (5.43) has two different expressions according to the relay position r and the correlation coefficient ρ of the symbols transmitted by the source and the relay. However, as we shall see, working with only one of them is enough for characterizing the OP behavior.

Theorem 5.2 (OP of DF)**.** *The probabilities involved in the OP of DF (5.43) can be found as follows. When $||r-d|| \neq D$ or $\rho \neq 0$ we have:*

$$\mathbb{P}\left(\mathcal{A}_{DF}^{c}(R,\rho) \cap \mathcal{B}_{DF}^{c}(R,\rho)\right) = \frac{\mu_2 \mathcal{L}_{I_d,I_r}\left(\frac{T(R)}{\mu_2},\frac{T(R)}{\mu_3}\right) - \mu_1 \mathcal{L}_{I_d,I_r}\left(\frac{T(R)}{\mu_1},\frac{T(R)}{\mu_3}\right)}{\mu_2-\mu_1}, \tag{5.44}$$

where:

$$\mu_1 = \frac{1}{2}\left[l_{sd}+l_{rd} - \left((l_{sd}-l_{rd})^2 + 4 l_{sd} l_{rd}|\rho|^2\right)^{\frac{1}{2}}\right], \tag{5.45}$$

$$\mu_2 = \frac{1}{2}\left[(l_{sd}+l_{rd}) + \left((l_{sd}-l_{rd})^2 + 4 l_{sd} l_{rd}|\rho|^2\right)^{\frac{1}{2}}\right], \tag{5.46}$$

$$\mu_3 = l_{sr}\left(1-|\rho|^2\right), \tag{5.47}$$

and $\mathcal{L}_{I_d,I_r}(\omega_1,\omega_2)$, the Laplace transform of the interferences, is given by (5.5), and $T(R)$ is given by (5.42). In addition, when $||r-d|| = D$ and $\rho = 0$, we have that $\mu_1 = \mu_2$ and:

$$\mathbb{P}\left(\mathcal{A}_{DF}^{c}(R,\rho) \cap \mathcal{B}_{DF}^{c}(R,\rho)\right) = \mathcal{L}_{I_d,I_r}\left(\frac{T(R)}{\mu_1},\frac{T(R)}{\mu_3}\right) - \frac{T}{\mu_1}\frac{d\mathcal{L}_{I_d,I_r}(\omega_1,T(R)/\mu_3)}{d\omega_1}\bigg|_{\omega_1=T(R)/\mu_1}. \tag{5.48}$$

Finally:

$$\mathbb{P}\left(\mathcal{A}_{DF} \cap \mathcal{A}_{DT}^{c}\right) = \mathcal{L}_{I_d}\left(\frac{T(R)}{l_{sd}}\right) - \mathcal{L}_{I_d,I_r}\left(\frac{T(R)}{l_{sd}}, \frac{T(R)}{\mu_3}\right). \tag{5.49}$$

Proof. See Appendix B.3. □

Notice that the OP depends only on the absolute value of ρ and not on its phase. This is a consequence of the uniform phase of the Rayleigh fading coefficients [33, 39]. The fact that the OP has two different expressions comes from the fading distribution that the destination sees from the joint source-relay transmission when $\rho = 0$ is different whether the source and relay are equidistant from the destination or not. However, the expression of the OP when $\rho = 0$ and $||r - d|| = D$, given by (5.48), can be obtained by continuously extending the other expression (5.44) at these points. Therefore, in what follows we shall focus our interest on (5.44), which fully characterizes the OP. Unfortunately, the OP cannot be evaluated to a simple expression, mainly because the integral in (5.8) does not have a closed form. However, it can be evaluated numerically without difficulty. Still, we can attempt to find the value of $|\rho|$ which minimizes the OP, by studying how this parameter appears in (5.44) and (5.49). Among other things, ρ allows some control over using DF or DT; this is because as $|\rho| \to 1$, $\mathcal{A}_{DF}$ given by (5.22) has a higher probability and hence, DF is used less. In the scenario in which transmissions are limited by independent background noises at the relay and the destination, there exists a value of $|\rho|$, generally non-zero, that minimizes the OP for each value of the system parameters [33, 38]. Although this is true for each realization of $\tilde{\Phi}$, if we average over the point process and the optimize the value of $|\rho|$ this may not be so. Unfortunately, a general analysis of the optimal $|\rho|$ is not possible, though we can consider the small node-density regime:

Lemma 5.3 (Optimal ρ)**.** *In the small node-density regime,*

$$\kappa(R) \leq \lim_{\lambda \to 0} \frac{\mathbb{P}\left(\mathcal{O}_{DF}(R, \rho, \lambda)\right)}{\lambda} = \lim_{\lambda \to 0} \frac{\mathbb{P}\left(\mathcal{O}_{DF}(R, 0, \lambda)\right)}{\lambda}.$$

This implies that, at least in the operating condition of the network, the OP when using DF is minimized by taking $\rho = 0$.

Proof. See Appendix B.4. □

As pointed out in [33] (see Remark 42) and [15], using $\rho = 0$ simplifies the implementation of DF because symbol synchronization between the source and its relay is not required. In what follows we focus in the small node-density regime and the simplified path loss function, and considering Lemma 5.3 we develop bounds on the OP taking $\rho = 0$. If numerical computation of the OP for other values is needed, it is straightforward to use Lemma 5.1 in Theorem

5.2. Notice that if we take $\rho = 0$, we have $\mathcal{B}_{DF} \subset \mathcal{A}_{DT}$ and hence, the protocol is always better than DT. It is straightforward to verify that in this case the OP can be bounded as:

$$P_{\text{out,DF}}(R,0) \leq \min\{\mathbb{P}(\mathcal{A}_{DF}(R,0)), \mathbb{P}(\mathcal{A}_{DT} \cap \mathcal{B}_{DF}(R,0)^c)\} + \mathbb{P}(\mathcal{B}_{DF}(R,0)) \tag{5.50}$$

$$= \min\{\mathbb{P}(\mathcal{A}_{DT}), \mathbb{P}(\mathcal{A}_{DT}(R,0)) + \mathbb{P}(\mathcal{B}_{DF}(R,0))\}. \tag{5.51}$$

This bound will be a good approximation, when the relay is close to source since a close inspection of $\mathcal{A}_{DF}$ and $\mathcal{B}_{DF}$ shows that in this setting the event $\mathcal{B}_{DF}$ will be dominant and $\mathcal{A}_{DF}$ will have a relatively small probability of occurrence. The following Theorem deals with evaluating the bound (5.51) and with bounding $\kappa(R)$ for the small node-density regime:

Theorem 5.4 (OP upper bounds for DF)**.** *When $||r-d|| \neq D$, considering the simplified path loss function, the OP for $\rho = 0$ can be upper bounded as:*

$$P_{out,DF}(R,0) \leq \min\Bigg\{1 - e^{-\lambda\delta D^2}, \Big(1 - e^{-\lambda\delta||r||^2}\Big) + \left(1 - \frac{D^\alpha e^{-\lambda\delta||r-d||^2} - ||r-d||^\alpha e^{-\lambda\delta D^2}}{D^\alpha - ||r-d||^\alpha}\right)\Bigg\}, \tag{5.52}$$

with δ given by (5.41). In addition, in the small node-density regime, $\kappa(R)$ can be bounded as:

$$\kappa_{DF}(R) \leq \delta \min\left\{D^2, \|r\|^2 + \|r-d\|^2 D^2 \frac{D^{\alpha-2} - \|r-d\|^{\alpha-2}}{D^\alpha - \|r-d\|^\alpha}\right\}. \tag{5.53}$$

Proof. See Appendix B.5. □

Notice again, that the case where $||r-d|| = D$ can be treated via continuity arguments as mentioned above. Finally, the following lower bound on the OP of half-duplex DF will be useful to compare it to the OP of full-duplex DF:

Theorem 5.5. *The OP of sequential DF can be lower bounded as:*

$$\mathbb{P}(\mathcal{O}_{SDF}(R,\varepsilon)) \geq 1 - \Bigg[\mathcal{L}_{I_d}\left(\frac{T(R)}{l_{sd}}\right) + \mathcal{L}_{I_d,I_r}\left(\frac{T(R)}{\tilde{\mu}_1}, \frac{2^{R/\varepsilon}-1}{l_{sr}}\right) + \frac{\tilde{\mu}_2 \mathcal{L}_{I_d,I_r}\left(\frac{T(R)}{\tilde{\mu}_2}, \frac{2^{R/\varepsilon}-1}{l_{sr}}\right) - \tilde{\mu}_1 \mathcal{L}_{I_d,I_r}\left(\frac{T(R)}{\tilde{\mu}_1}, \frac{2^{R/\varepsilon}-1}{l_{sr}}\right)}{\tilde{\mu}_2 - \tilde{\mu}_1}\Bigg], \tag{5.54}$$

with $T(R)$ given by (5.42) and:

$$\tilde{\mu}_1 = \frac{1}{2}\left[l_{sd} + (1-\varepsilon)l_{rd} - |l_{sd} - (1-\varepsilon)l_{rd}|\right], \tag{5.55}$$

$$\tilde{\mu}_2 = \frac{1}{2}\left[l_{sd} + (1-\varepsilon)l_{rd} + |l_{sd} - (1-\varepsilon)l_{rd}|\right]. \tag{5.56}$$

Proof. See Appendix B.6. □

5.3.2 Compress-and-forward

In this section we derive an upper bound for the OP of CF. This analysis is far more involved than that of DF because, as we mentioned earlier, the condition of full decoding that is imposed on the relay for DF results in an achievable rate which does not depend on the spatial correlation of the interference signals at the relay and at the destination. Since in CF the relay searches for a sequence which acts as a compressed version of what it receives, without decoding the message from the source, the correlation of the interference signals given by (5.4) does affect the achievable rate. In [40], the authors carry out an analysis of Gaussian relay channels with correlated noises in which the correlation coefficient is fixed and full CSI is available at the relay. This implies that the compression variance n_c can be chosen to achieve equality in (5.35). Under these conditions, the authors compare the performance of CF with correlated and uncorrelated noises and show that negative noise correlation always helps CF, while positive correlation sometimes helps CF. In the setup of this chapter, since full CSI is not available at the relay, the compression variance has to be chosen a priori and the additional outage event that (5.35) is not met has to be considered. In addition, it is straightforward to show that under this condition, in which the variance n_c is a fixed constant independent of network parameters, the value of ρ_N which maximizes or minimizes the rate could be located anywhere on the disc $|\rho_N| \leq 1$. Since no closed-form analysis can be carried out considering this random correlation, we resort to a procedure which allows us to bound the effect of the correlation of the interference on the achievable rate.

Lemma 5.6 (Rate gap in CF)**.** *The achievable rate of CF for any spatial noise correlation ρ_N is at most one bit worse than the rate with uncorrelated noises, that is:*

$$R_{CF}(\rho_N, n_c) \geq R_{CF}(0, n_c) - 1. \tag{5.57}$$

Proof. See Appendix B.7. □

Using the previous lemma we can work with the OP of CF assuming that the interference time signals are uncorrelated and increase the attempted rate to bound the actual value of the OP, which leads to the following result:

Theorem 5.7 (OP bound for CF)**.** *For an attemped rate R the OP of CF can be upper bounded as:*

$$\begin{aligned} P_{out,CF}(R, n_c) &\leq \mathbb{P}(\mathcal{O}_{CF}(R+1, n_c, 0)) \\ &= \mathbb{P}(\mathcal{A}_{CF}(R+1, 0) \cup \mathcal{B}_{CF}(n_c, 0)). \end{aligned} \tag{5.58}$$

In addition we can bound the OP with $\rho_N = 0$ as:

$$\mathbb{P}(\mathcal{A}_{CF}(R,0)) \leq 1 - e^{-\frac{T(R)n_c}{l_{sr}}}\mathcal{L}_{I_r}\left(\frac{T(R)}{l_{sr}}\right) - \left[\sum_{\tilde{n}=0}^{\tilde{N}-1} e^{\frac{\tilde{n}n_cT(R)}{\tilde{N}l_{sr}}}\mathcal{L}_{I_d,I_r}\left(\frac{(\tilde{N}-\tilde{n})T(R)}{\tilde{N}l_{sd}}, \frac{\tilde{n}T(R)}{\tilde{N}l_{sr}}\right)\right.$$
$$\left. - e^{\frac{(\tilde{n}+1)n_cT(R)}{\tilde{N}l_{sr}}}\mathcal{L}_{I_d,I_r}\left(\frac{(\tilde{N}-\tilde{n})T(R)}{\tilde{N}l_{sd}}, \frac{(\tilde{n}+1)T(R)}{\tilde{N}l_{sr}}\right)\right], \quad (5.59)$$

and:

$$\mathbb{P}(\bar{\mathcal{A}}_{CF}(R,0) \cap \mathcal{B}_{CF}(n_c,0)) \leq$$
$$1 - \mathbb{E}\left[\mathcal{L}_{I_d,I_r}\left(\frac{(1+T(R))l_{sr}|h_{sr}|^2}{T(R)n_cl_{rd}}, \frac{(1+T(R))l_{sd}|h_{sd}|^2}{T(R)n_cl_{rd}}\right)\right]. \quad (5.60)$$

The expectation is over h_{sr} and h_{sd}, with $T(R)$ given by (5.42). For the simplified path loss we can bound (5.60) to obtain:

$$\mathbb{P}(\bar{\mathcal{A}}_{CF}(R+1,0) \cap \mathcal{B}_{CF}(n_c,0)) \leq 1 -$$
$$\mathbb{E}\left[\mathcal{L}_{I_d}\left(\frac{(1+T(R))l_{sr}|h_{sr}|^2}{T(R)n_cl_{rd}}\right)\right]\mathbb{E}\left[\mathcal{L}_{I_r}\left(\frac{(1+T(R))l_{sd}|h_{sd}|^2}{T(R)n_cl_{rd}}\right)\right]. \quad (5.61)$$

Proof. See Appendix B.8. □

The previous bound has the advantage of allowing us to upper bound the OP of CF in a rigorous manner by considering the gap with the worst case of rate that may be achieved when the spatial correlation between the time signals is the worst possible. Since the correlation of the interference plays a role in the achievable rate it is possible that the bound is conservative of the performance of CF.

5.4 Numerical results

In this section we present figures to study the behavior of the derived expressions and to compare the performance of DF and CF with DT and half-duplex DF. In all our simulations we take the destination at $d = (10,\ 0)$ and the path loss exponent $\alpha = 4$.

In Fig. 5.5 we can see the comparison of DF versus DT, both through the exact numerical evaluation of the OP using Theorem 5.2 and Lemma 5.1, and with the upper bounds given by (5.52), for different relay positions, taking $\rho = 0$ and $R = 0.5$ b/use. We can see that these bounds are accurate when the OP is small, and the relay is close to the source, as proposed. In addition, for a fixed source-relay distance the OP increases as the relay grows further away from the destination. This is because the probability of $\mathcal{B}_{DF}(R,\rho)$ increases as this happens.

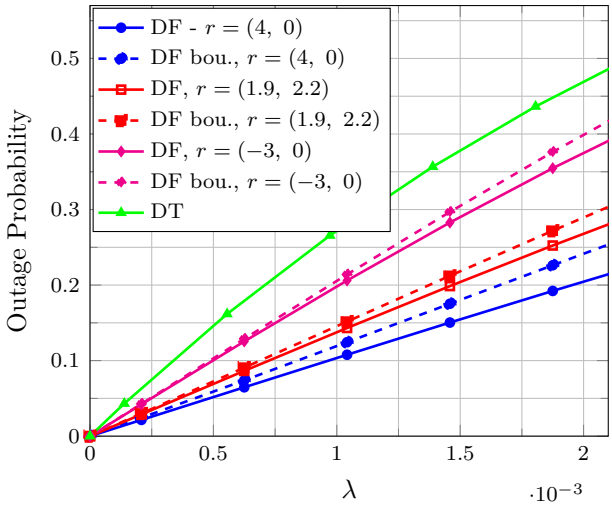

Figure 5.5: OP of DF with the relay located at r versus DT. Exact expressions for DF come from using Theorem 5.2 and Lemma 5.1 while upper bounds come from (5.52). $d = (10, 0)$, $R = 0.5$ b/use, $\alpha = 4$, $\rho = 0$.

In Fig. 5.6 we can observe how the variation of the true OP of DF given by (5.43) and Theorem 5.2, as a function of the correlation between the symbols of the source and the relay $|\rho|$, for various relay positions, for an attempted rate $R = 1$b/use. Two sets of curves are presented. One for the case of $\lambda = 10^{-4}$, in which the OP is small, and the other for $\lambda = 10^{-3}$, in which the OP is larger. In both cases we see, as Lemma 5.3 states, $\rho = 0$ is the optimal choice. In Fig. 5.7 we compare the maximum rate R that can be attempted, given a maximum allowed OP of 0.05, when the relay is located on the line between the source and destination. We consider full-duplex DF (Theorem 5.2), half-duplex DF (Theorem 5.5), CF (Theorem 5.7), and DT (Eq. (5.40)). In the case of half-duplex DF we numerically optimize ε, the fraction of the block in which the relay listens and attempts to decode. The same is done for the compression variance n_c of CF. It can be seen that full-duplex DF outperforms half-duplex DF, specially when the relay is equidistant between the source and destination. The gains however, are not as large near the source or the destination. However, half-duplex DF requires the optimization of ε which, as the relay moves away from the source and closer to the destination, takes values on the whole interval $(0, 1)$, while for the full-duplex version it suffices to take $\rho = 0$. On the other hand, we see that for CF, the bounds do not predict that CF is better than DF when the relay is near the source, as was observed in other scenarios. This hints that the correlation of the interference may have an important effect on the performance of CF. In order to explore this, in Fig. 5.8 we plot the spatial regions in which DF or CF are preferred over the other, for $R = 2$b/use and $\alpha = 4$. The performance

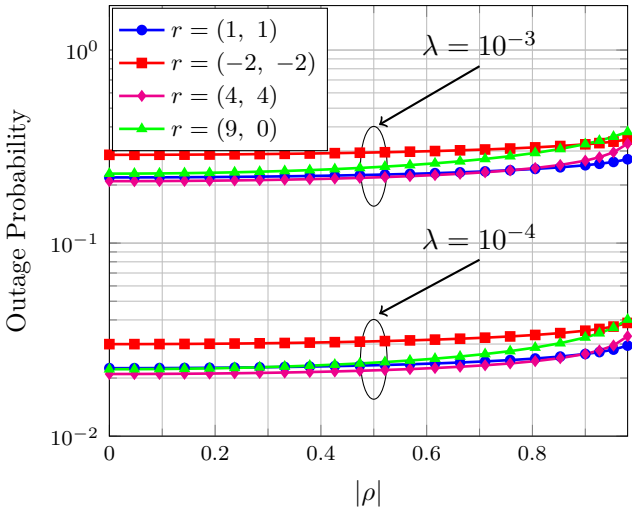

Figure 5.6: Outage probability as a function of $|\rho|$ for various relay positions and for $\lambda = 10^{-3}$ and $\lambda = 10^{-4}$. $d = (10,\ 0)$, $\alpha = 4$ and $R = 1$b/use. The OP is found by using Theorem 5.2 and Lemma 5.1.

of CF is estimated by performing a Monte Carlo simulation of the point process, optimizing the noise variance n_c and estimating the OP. We see that, as expected, CF performs better than DF when the relay is closer to the destination, while DF is better everywhere else. This shows that, in fact, CF can take advantage of the interference correlation, which impacts its performance significantly, which DF cannot. The optimization of the variance n_c, however, is a difficult problem which is not present in full-duplex DF. As mentioned before, DT is always worse than the two other protocols, which are both close to DT when the relay is very far from both the source and the destination. In the plotted region, however, DT was not yet close to DF or CF.

5.5 Summary and final remarks

In this chapter we have analyzed the performance, in terms of the OP, of a relay channel employing DF and CF when the interference comes from a network in which nodes are distributed as a PPP. We have derived an expression for the OP of full-duplex DF and upper bounds on the OP which are amenable for analysis and tight when the OP is small. We have also determined the correlation coefficient ρ of the symbols transmitted by the source and the relay which minimizes the OP as the density of interferers vanishes. We showed that the same analysis cannot be carried out for CF, because the achievable rates with this protocol are dependent on the spatial correlation of the interference signals. To avoid this issue, we

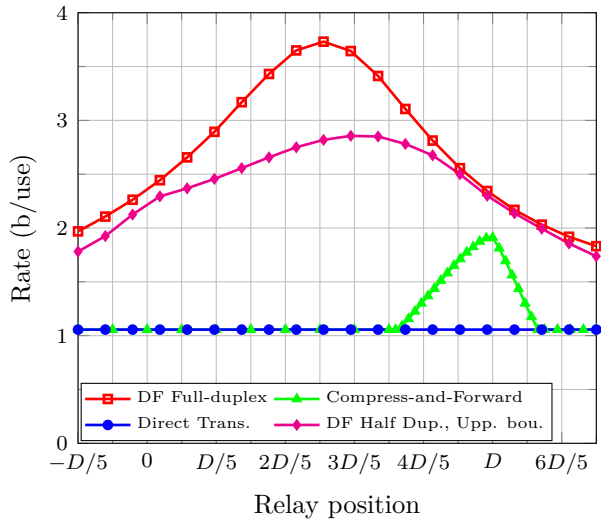

Figure 5.7: Maximum rate achievable through all the studied protocols when the relay is aligned with the source and destination and an OP smaller that 0.05 is required. The OP of full-duplex DF comes from Theorem 5.2, half-duplex DF from Theorem 5.5, CF from Theorem 5.7 and DT from (5.40). The Laplace transforms are numerical from Lemma 5.1. $d = (10, 0)$, $\alpha = 4$.

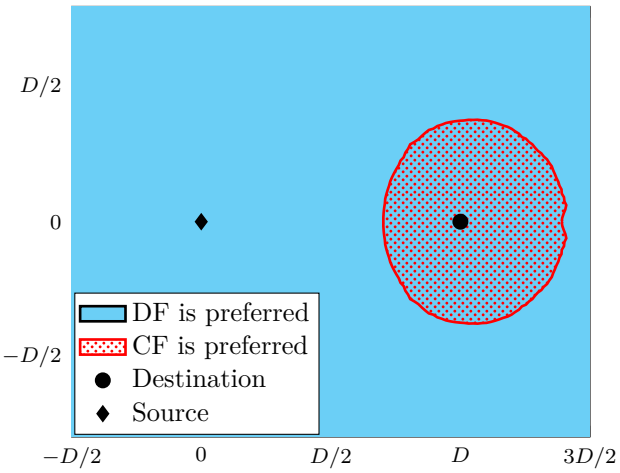

Figure 5.8: Spatial positions in which each scheme (DF or CF) is preferred. The OP of DF comes from Theorem 5.2 with $\rho = 0$, and the performance of CF is estimated by performing a Monte Carlo simulation of the point process, optimizing the noise variance n_c and estimating the OP. DT is not shown because it is not better than the other protocols in the plotted region. $\lambda = 0.5 \times 10^{-4}$. $d = (10, 0)$, $R = 4$b/use, $\alpha = 4$.

showed that for any spatial correlation of the point process, the rate achievable is at most one bit worse than the rate when the interferences are uncorrelated. We compared the OP of both full-duplex protocols with half-duplex DF and DT. We have observed that full-duplex DF is the best alternative over the other protocols, except near the destination, where CF is better. However, both CF and half-duplex DF, require the optimization of parameters which depend strongly on the relay's position, while full-duplex DF does not. The disadvantage of the full-duplex protocols comes from the practical aspects involving simultaneous transmission and reception. This analysis may serve as a starting point for the analysis of more complex network systems and to study the interactions among nodes in large wireless networks, involving different coding and medium-access schemes and network infrastructure.

Chapter 6

Analysis of a Cooperative Strategy for a Large Decentralized Wireless Network

Summary. This chapter investigates the benefits of cooperation and proposes a relay activation strategy for a large wireless network with multiple transmitters. In this framework, some nodes cooperate with a nearby node that acts as a relay, using the DF protocol, and others use DT. The network is modeled as an independently marked PPP and the source nodes may choose their relays from the set of inactive nodes. Although cooperation can potentially lead to significant improvements in the performance of a communication pair, relaying causes additional interference in the network, increasing the average noise that other nodes see. We investigate how source nodes should balance cooperation vs. interference to obtain reliable transmissions, and for this purpose we study and optimize a relay activation strategy with respect to the OP. Surprisingly, in the high reliability regime, the optimized strategy consists on the activation of all the relays or none at all, depending on network parameters. We provide a simple closed-form expression that indicates when the relays should be active, and we introduce closed form expressions that quantify the performance gains of this scheme with respect to a network that only uses DT.

6.1 Introduction

Cooperative wireless networks in which relay nodes can be used to increase throughput and reliability have been studied in the past [33]. Although the capacity of the single-relay channel [14] remains unsolved and its optimal coding scheme unknown, there have been significant advances in quantifying the performance gain obtained through cooperation. However, finding capacity regions or analyzing the performance of large random wireless networks may be, if feasible, very hard. As an alternative, spatial models employing tools from stochastic geometry and graph theory provide a comprehensive framework for the analysis of large wireless networks [10, 42].

As we have seen in the previous chapter, the OP is a useful metric to study wireless

networks [10, 42, 52] in which the users are assumed to be unaware of the instantaneous parameters of the network and cannot optimize their behavior to attain successful transmissions. Among other reasons, the relevance of the OP comes from the fact that, in an outage event, sent messages cannot be successfully transmitted, and hence the overall delay of the network is increased due to retransmissions. In the previous chapter, we studied the performance, in terms of the OP, of a single relay channel in the presence of interference generated by a spatial configuration of nodes modeled through a point process. While this is a best case scenario from the point of view of the source having a relay, it is reasonable to assume that other sources will also be interested in one. For this reason, in this chapter we focus on analyzing the performance of a large decentralized wireless network in which transmitters may be aided by nearby relays. More precisely, we consider a network formed by two types of clusters: source-relay-destination clusters, which use the full-duplex DF [14] scheme, and clusters with source-destination pairs which employ simple DT. These clusters could be interpreted as a single hop in a multi-hop transmission scheme or by themselves as single-hop communications. One of the central motivations behind this analysis is to provide an understanding of the limitations and benefits of cooperation in such decentralized scenarios. In fact, the advantage of cooperation among nodes for an individual source-destination link was widely studied in the past years, addressing both theoretical and practical issues [15, 16, 33]. In this chapter we analyze a scenario in which the communication impairments are caused by a network of users which are also attempting to achieve successful transmissions through cooperation and cause interference to each other. It is clear that relays can significantly improve the rate and reliability of a single source-destination pair. However, in a large wireless network, the nodes will observe an increase in their interference levels as more relays are activated. This means that while cooperation may be beneficial locally, globally its benefits may be reduced.

As in Chapter 5, the network is modeled as an independently marked homogeneous PPP (Sec. 4.3), limited by the SIR, where signal attenuation occurs both through path loss and slow fading. We focus directly on the OP and specifically in the high reliability regime [52] we described in Sec. 5.2.2, and the spatial contention parameter introduced in Eq. (5.18).

The transmission scheme of the network is a mixed one, since some clusters will be using the DF scheme while others will employ DT (see Fig. 6.1). It is assumed that almost no CSI is available at the transmitting nodes, which is often the case in decentralized wireless networks without feedback. Only a rough estimation of the position of nearby potential relays may be available, and hence, it can be used for the relay selection. We assume that each source chooses its potential relay among the nodes that are not transmitting as its nearest neighbor on a cone with aperture angle ϕ_0, centered toward its destination (see Fig. 6.2). This scheme will increase the likelihood of finding a relay which is close to the source and at the same

time reduces the effect of the path loss on the relay-destination link. Notice that this effect is minimized if the relay, source and destination are aligned. As a special case of this scenario, the relay can be chosen as the nearest neighbor of the source on the whole plane, requiring the least amount of CSI. The motivation behind choosing the nearest neighbor as a relay comes from the fact that DF is nearly optimal from an information theoretic point of view [14, 33], and as seen in Chapter 5, when the relay is not far from the source. In this case, the probability of the relay not being able to decode the source message is minimized. A simple random relay activation scheme is introduced in which each candidate relay node decides whether to be active or not in a random manner, independently of each other, and of all network parameters. This simple activation scheme will act as a means of controlling the relay density in the network while still retaining a balance between interference generation and cooperation.

6.1.1 Related work

Over the past years, the performance gains of cooperative communications in relay networks were widely studied from an information-theoretic perspective. Since the seminal work of Cover and El Gamal [14], several contributions have been published on the subject. More recently, the emphasis has been put on studying the performance of wireless relay channels where outage performance and ergodic rates of fading channels with Gaussian noise have been derived (see [15, 16, 33, 38] and the references therein). Among these valuable studies, the only impairments to the communication were due to additive Gaussian noise and fading, and very little attention was paid to the effect of the interference generated (or suffered) by other users. However, interference is the major impairment in many wireless networks.

The study of the capacity of general wireless networks taking into account the interference generated by the different users was pioneered by the seminal work of Gupta-Kumar [62], where the concept of transport capacity and fundamental scaling laws on the network throughput were obtained considering only point-to-point coding. In [63] multiuser achievability regions were obtained and it was shown that for some special wireless networks significantly better scaling laws on the network throughput, with respect to the case in [62], are possible. Further progress was done in [64], where new scaling laws were derived using coherent multistage relaying with interference subtraction and in [65], where extensions to fading channels were obtained.

Stochastic geometry and point processes (see Chapter 4) are not only elegant mathematical frameworks but also useful tools to deal with more realistic network models, where the spatial position of nodes and the effect of interference can be incorporated in a probabilistic manner. Although several types of point processes can be used to model different kind of networks,

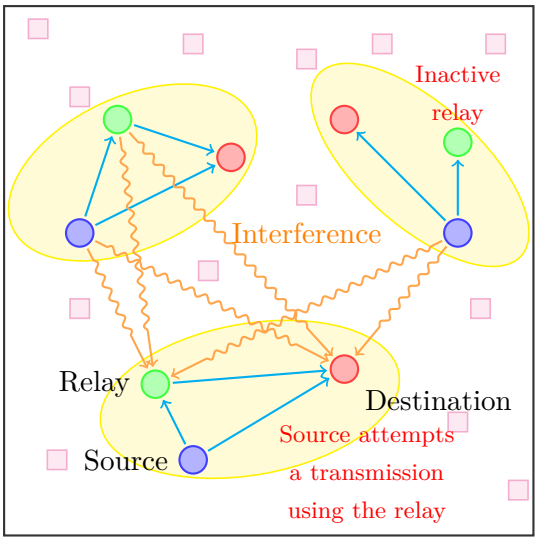

Figure 6.1: The network is formed of clusters (in yellow) employing decode-and-forward or direct transmission. Relays (in green) are chosen from the set of inactive users (squares in magenta) to aid the sources (in blue) to transmit their messages to their destinations (in red).

it is the homogeneous PPP which has received the most attention. Although other types of point processes can provide more realistic models in certain cases [52, 66, 67], the extended use of the homogeneous PPP comes from the possibility to obtain simple closed-form results in several cases of interest. Initial works focus specially in single-antenna point-to-point links [10, 68, 69]. There are also other works in which more complex scenarios are considered, such as broadcasting with superposition coding [70], simultaneous transmissions to a single user [71], or multi-antenna strategies [50].

6.1.2 Contributions

The main contribution of this chapter is studying and optimizing the network management strategy for activating the relays in the proposed transmission scheme. The main conclusion is that in the high reliability regime this optimized scheme consists on turning all the relays on or off simultaneously, that is, the optimal relay activation probability is either 0 or 1. To do this, we develop closed-form approximations for the OP of the network, and study the interference-cooperation balance by finding the relay activation probability which minimizes the average OP. Moreover, the network parameter regions in which all the relays should be on are identified, and a simple relay activation scheme which is close to the optimal behavior is introduced. Finally we provide simple expressions that quantify the performance gains in terms of OP for the scheme with the optimal relay activation scheme with respect to a

network in which all users employ DT.

The chapter is organized as follows. In Section 6.2, a general and a mathematical descriptions of the network model are presented. We also discuss the DF scheme and its achievable rate in the assumed network model. In Section 6.3 we introduce an expression for the OP for this network, deriving closed form approximations to it. In Section 6.4 we study the performance of the network, finding the optimal relay activation probability, identifying the network parameters for which all the relays should be on or off, introducing the relay activation policy and comparing the performance of this scheme against DT. In Section 6.5 we present some numerical simulations and in Section 6.6 we provide some concluding remarks. Finally, the mathematical proofs are deferred to Appendix C.

Notation

In this chapter we use the big O notation: $f(x) = O\left(g(x)\right)$ as $x \to x_0$ if there exists $M > 0$ such that $|f(x)| \leq M|g(x)|$ in some neighborhood of x_0.

6.2 General considerations and network model

6.2.1 The model

We consider a spatial network model in $\mathbb{R}^2$ in which source nodes generate messages and attempt to transmit them to intended destinations, either through a direct link, in which case the destinations receive symbols from their sources only, or by using others nodes as relays. Every relay aids a single source node, acting as a secondary full-duplex transmitter sharing the same time slots and frequency band. This setup allows the nodes to be grouped into *clusters* formed by a source-destination pair or by a source-relay-destination triplet, if the source has an associated relay, as shown in Fig. 6.1.

We start from a set of nodes Φ which we assume forms an homogeneous PPP of density λ. Some nodes from this set choose to access the network and become sources using slotted ALOHA [10] with transmit probability λ_s/λ. This splits the set Φ into two new independent homogeneous PPPs:

- Φ_s of sources of density λ_s,
- Φ_{in} of potential relay nodes of density $\lambda_{in} = \lambda - \lambda_s$,

such that $\Phi = \Phi_s \cup \Phi_{in}$. Notice that the proportion of sources and potential relays can be adjusted by the medium access probability.

Inactive nodes should then be assigned in a one-to-one fashion to each source such that cooperation is beneficial. To simplify the relay assignment strategy we shall assume that the spatial density of the sources is much smaller than that of potential relays, i.e. $\lambda_s \ll \lambda_{in}$. Under this hypothesis, we will neglect the probability of two sources choosing the same inactive node as a relay, since each source will have a rich selection of relay candidates in its vicinity ([68], Ex. 3). Thus we can simplify our model by including the position of the potential relay and its activation scheme as an independent mark to each source, obtaining the spatial distribution of the relay from the original homogeneous PPP Φ_{in} of intensity λ_{in} to which the relays are assumed to belong.

We consider the usual and realistic assumption that only little or no CSI is available, while nodes may have some estimation of the spatial position of neighboring nodes. For this reason nodes cannot adjust their rates to achieve a reliable communication according to instantaneous conditions, but may use this spatial knowledge to select a relay.

Based on these considerations, the network is modeled as an independently marked PPP (Section 4.3):

$$\tilde{\Phi}_s = \left\{ (x_i, (\varepsilon_{x_i}, k_{x_i}, \theta_{x_i}), h_{x_i r}, h_{x_i d}, h_{k_{x_i} r}, h_{k_{x_i} d}) \right\}, \tag{6.1}$$

such that:

- The positions of the sources form the homogeneous PPP $\Phi_s = \{x_i\}$ of intensity λ_s.
- The triplet $(\varepsilon_x, k_x, \theta_x)$ models the relay position and its state. The random variable θ_x, uniform in $[0, 2\pi)$, models the direction of each destination relative to its source, with $\theta = 0$ meaning that the destination is in the direction of the canonical vector $(1, 0)$ with respect to its source. k_x indicates the position of the potential relay relative to its source, that is, the potential relay for source x is located at $x + k_x$. According to what we mentioned earlier, the relay will be chosen as the nearest neighbor of the source on a cone of aperture ϕ_0 with the destination on its axis. This means that the distribution of the potential relay k_x for a source at the origin, conditioned on the direction of the destination θ_x will be (in polar coordinates) [72]:

$$f_{k_x|\theta}(\rho, \phi) = \lambda_{in} \rho e^{-\lambda_{in} \phi_0 \rho^2/2} \mathbb{1}\{|\phi - \theta| < \phi_0/2\} \times \mathbb{1}\{0 \leq \phi \leq 2\pi, \rho \geq 0\}. \tag{6.2}$$

  Using $\phi_0 = 2\pi$ means choosing the relay as the nearest neighbor on the whole plane instead of a cone, independently of the direction of the destination and using the least CSI. Notice that in this case the distribution of the nearest neighbor (6.2) becomes a bidimensional Gaussian RV of variance:

$$\sigma_{in}^2 = \frac{1}{2\pi\lambda_{in}}. \tag{6.3}$$

Notice that we can parameterize the nearest neighbor distribution (6.2) in terms of σ_{in} and for any cone aperture ϕ_0 as:

$$f_{k_x|\theta}(\rho,\phi) = \frac{1}{2\sigma_{in}^2}\rho e^{-\frac{\phi_0\rho^2}{4\pi\sigma_{in}^2}} \mathbb{1}\{|\phi-\theta| < \phi_0/2\}\mathbb{1}\{0 \le \phi \le 2\pi, \rho \ge 0\}. \tag{6.4}$$

This implies that the effect of considering the nearest neighbor on a cone is simply restricting the nearest neighbor on the plane distribution (Gaussian) to the cone and increasing the variance (by means of the ϕ_0 in the exponent). Thus, we can study the relay activation strategy in terms of the variance of the nearest neighbor on the whole plane (σ_{in}) and cone aperture ϕ_0 that the source uses. Additionally, notice that reducing the cone aperture allows the relay to be located towards the destination but at the same time, the increased variance implies that the relay will be, on average, farther from the source than if we take $\phi_0 = 2\pi$.

The random variable ε_x indicates if the corresponding source uses a relay or not. In our case, we take it to be a Bernoulli random variable with success probability p_r, independent of everything else. Notice that the parameter p_r allows the adjustment of the relay density and hence allows to control the additional interference introduced in the network, weighing the local and global effects of cooperation. In addition to a medium access scheme, this parameter can also be used to model the unavailability of a relay for reasons which are out of the control of the relay itself (such as a malfunction or a depleted battery). In such a case, the independent occurrence of these events among the relays is a reasonable assumption.

- All nodes transmit with unit power while the power received at y by a transmitter located at x is $|h_{xy}|^2 l_{xy}$ where $l_{xy} = \|x-y\|^{-\alpha}$ ($\alpha > 2$) is the usual path loss function and $|h_{xy}|^2$ is the power gain of Rayleigh fading with unit mean.

- An additional source with the same marks as the others, independent of the point process $\tilde{\Phi}_s$ and with its destination at $d = (D, 0)$, is added at the origin. The position of the relay for this source node will be r (with the same distribution as the $\{k_{x_i}\}_i$ random variables). The coefficients $|h_{sr}|^2$, $|h_{rd}|^2$ and $|h_{sd}|^2$ model the source-relay, relay-destination, and source-destination fading coefficients of this cluster, respectively. Theorem 4.6 (Slyvniak-Mecke [28, 8]) and the stationarity of the process guarantee that the study of this cluster's behavior will be representative of the behavior of any other cluster in the network.

- h_{xr} and $h_{k_x r}$ model the fading gains from each source and its relay to the relay of the source at the origin, while h_{xd}, $h_{k_x d}$ model the gains from each source and its relay to the destination of the source at the origin.

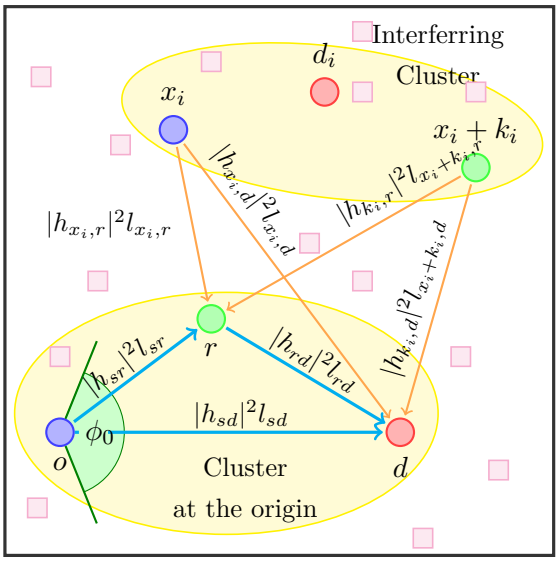

Figure 6.2: The relay is chosen as the nearest neighbor of the source on a cone of aperture ϕ_0 with its axis aligned with the destination. Also the power fading within the cluster at the origin and for the interference from other clusters is shown. Magenta squares represent inactive users.

Please see Fig. 6.2 for a graphical representation of the key parameters in the model.

Remark 6.2.1. *Notice that other schemes for activating and selecting the relays based on position can be studied by appropriately selecting the triplet* $(\varepsilon_x, k_x, \theta_x)$. *For example, we could study the performance of choosing the nearest or the farthest neighbor on a cone of finite radius as a relay. The probability of activating a relay will be that of finding at least one potential relay in the cone and the conditional distributions of the selected relay, given that the cone is not empty, can be found in [73].*

In the same way as in Chapter 5, each source and its relay (if it is active) use unit-variance Gaussian signaling and their codebooks have correlation coefficient ρ. In addition, the codebooks between different clusters are independent. Destination and relay nodes in each cluster attempt to decode their messages while treating the interference from other clusters as noise. With these hypotheses the aggregate interferences at the relay and destination of the typical cluster are zero-mean CCSG variables whose variances, conditioned on $\tilde{\Phi}_s$, are given by:

$$I_r = \sum_{x \in \Phi_s} \left[\frac{|h_{xr}|^2}{\|x - r\|^\alpha} + \varepsilon_x \left(\frac{|h_{k_x r}|^2}{\|x + k_x - r\|^\alpha} + \frac{2\Re\left\{ h_{xr} h_{k_x r}^* \rho \right\}}{\|x - r\|^{\frac{\alpha}{2}} \|x + k_x - r\|^{\frac{\alpha}{2}}} \right) \right], \tag{6.5}$$

$$I_d = \sum_{x\in\Phi_s}\left[\frac{|h_{xd}|^2}{\|x-d\|^\alpha} + \varepsilon_x\left(\frac{|h_{k_xd}|^2}{\|x+k_x-d\|^\alpha} + \frac{2\Re\left\{h_{xd}h_{k_xd}^*\rho\right\}}{\|x-d\|^{\frac{\alpha}{2}}\|x+k_x-d\|^{\frac{\alpha}{2}}}\right)\right]. \tag{6.6}$$

6.2.2 Achievable rates and outage events

We have already discussed the main coding strategies for the relay channel, including DF, in Section 5.2.3 so we will not go into all the details once more here. It suffices to say that using DF with *regular encoding* and *sliding-window decoding* the following rate is achievable:

$$R_{DF} = \max_{\rho\in\mathbb{C},|\rho|\leq 1}\min\left\{\mathcal{C}\left(\frac{|h_{sr}|^2 l_{sr}(1-|\rho|^2)}{I_r}\right),\right.$$
$$\left.\mathcal{C}\left(\frac{|h_{sd}|^2 l_{sd} + |h_{rd}|^2 l_{rd} + 2\sqrt{l_{sd}l_{rd}}\Re\left(\rho h_{sd}h_{rd}^*\right)}{I_d}\right)\right\}, \tag{6.7}$$

where $\mathcal{C}(u) = \log_2(1+u)$. Also, as we mentioned in Section 5.2.3, this protocol is *oblivious* [38] to the presence of the relay, that is, the source can use the same coding scheme for DF or DT without considering if the relay is present or not. This is very important, since the relay can decide to activate itself (achieving the DF rate) or not (achieving the DT rate) without taking into account the source, which in both cases employs the same coding scheme. The interferences powers are now given by (6.5) and (6.6), in which some of the sources generating the interference may also have relays. This also implies that the interference powers also depend on the correlation coefficient ρ between the codebooks used by each source and its relay. In the previous chapter we analyzed the value of ρ that minimizes the OP in the high reliability regime and found that, when only the source has a relay, the optimal value is $\rho = 0$ (Lemma 5.3). For this reason, and because it simplifies the implementation of DF ([33, Rem. 42],[15]), in this chapter we consider the case $\rho = 0$[1]. The outage events when the relay is present are the same as those in the previous chapter. The event that the relay does not decode $\mathcal{A}_{DF}(R,\rho)$ is given by (5.22), which, taking $\rho = 0$ is:

$$\mathcal{A}_{DF}(R) \triangleq \mathcal{A}_{DF}(R,0) = \left\{|h_{sr}|^2 l_{sr} < T(R)I_r\right\}, \tag{6.8}$$

with $T(R) = 2^R - 1$ as in the previous chapter, R being the attempted rate. The event that the destination cannot decode the joint transmission from the relay and the source is $\mathcal{B}_{DF}(R,\rho)$, given by (5.26), which taking $\rho = 0$, becomes:

$$\mathcal{B}_{DF}(R) \triangleq \mathcal{B}_{DF}(R,0) = \left\{|h_{sd}|^2 l_{sd} + |h_{rd}|^2 l_{rd} < TI_d\right\}. \tag{6.9}$$

[1]It may be possible to prove, following the lines of the proof of Lemma 5.3, that for this network the optimal value in the high reliability regime is also $\rho = 0$.

The outage event when the relay is present is therefore $\mathcal{A}_{DF}(R) \cup \mathcal{B}_{DF}(R)$. When there is no relay and the sources uses DT, the outage event is (5.27), and the OP is the same as (5.40):

$$P_{\text{out,DT}} = 1 - \exp\{-\lambda\delta D^2\}, \tag{6.10}$$

with δ given by (5.41) and $\tilde{C}$ given by (5.10).

Using the asymptotic expansion of the OP we can write

$$P_{\text{out,DT}}(R) = \lambda_s \delta D^2 + O((\lambda_s \delta D^2)^2)$$

as $(\lambda_s \delta D^2)^2 \to 0$. In the high reliability regime, when the success probability of the network is close to one, a reasonable approximation is to neglect the higher order $O(\cdot)$ term and write $P_{\text{out,DT}}(R) \approx \lambda_s \delta D^2$, meaning that the approximation will be good and that $\lambda_s \delta D^2$ will be small. In this expansion we see that $\gamma = \delta D^2$ is the contention parameter of the network, as defined in Section 5.2.2. As we shall see, the activation or not of the relay will not depend only on the density of interferers, which is why we consider an expansion in terms of $\lambda_s \delta D^2$ instead of λ_s only.

6.3 The outage probability of the network

In this section we study the OP of the network as introduced in the previous section. By conditioning on the fact that the cluster at the origin uses a relay or not, and on this relay position, we can see that the OP of the cluster at the origin (and hence of any other cluster) can be written as:

$$\begin{aligned} P_{\text{out,mix}}(R) = \mathbb{P}\left(\varepsilon_0 = 0\right) \mathbb{P}\left(\mathcal{A}_{DT}(R) | \varepsilon_0 = 0\right) \\ + \mathbb{P}\left(\varepsilon_0 = 1\right) \mathbb{E}_r \left[\mathbb{P}\left(\mathcal{A}_{DF}(R) \cup \mathcal{B}_{DF}(R) | r, \varepsilon_0 = 1\right)\right]. \end{aligned} \tag{6.11}$$

This expression can be evaluated in terms of the Laplace transform of the interference random variables I_r and I_d, as the following theorem states:

Theorem 6.1. *The outage probability of the network $P_{out,mix}$ given by (6.11) can be written as:*

$$\begin{aligned} P_{out,mix}(R) = \mathbb{P}(\varepsilon_0 = 0) \left[1 - \mathcal{L}_{I_d}\left(\frac{T(R)}{l_{sd}}\right)\right] + \mathbb{P}(\varepsilon_0 = 1) \\ \times \mathbb{E}_r \left[\frac{D^\alpha \mathcal{L}_{I_d,I_r}\left(\frac{T(R)}{l_{rd}}, \frac{T(R)}{l_{sr}}\right) - ||r-d||^\alpha \mathcal{L}_{I_d,I_r}\left(\frac{T(R)}{l_{sd}}, \frac{T(R)}{l_{sr}}\right)}{D^\alpha - ||r-d||^\alpha}\right], \end{aligned} \tag{6.12}$$

where $\mathcal{L}_{I_d,I_r}(\cdot,\cdot)$ is the Laplace transform of the interferences, whose definition is given by (5.5). Additionally, setting $\omega_2 = 0$ in (5.5) we obtain $\mathcal{L}_{I_d}(\omega_1)$, the Laplace transform of the interference at the destination.

Proof. It is identical to the proof of Theorem 5.2 taking $\rho = 0$, with the exception that the Laplace transforms are with respect to the interference powers given by (6.5) and (6.6). □

Since the Laplace transform of the interferences can be interpreted as a functional of the marked point process $\tilde{\Phi}$, we can use (4.14) to evaluate this transform as:

$$\mathcal{L}_{I_d,I_r}(\omega_1,\omega_2) = \exp\left\{-\lambda_s p_r t(\omega_1,\omega_2,r,d)\right\} \times \exp\left\{-\lambda_s(1-p_r)\left[C(\omega_1^{2/\alpha}+\omega_2^{2/\alpha}) + f(\omega_1,\omega_2)\right]\right\}, \quad (6.13)$$

where f comes from (5.11),

$$t(\omega_1,\omega_2,r,d) = \int_{\mathbb{R}^2}\int_{\mathbb{R}^2}\left[1 - z\left(\omega_1,x,k,d\right) z\left(\omega_2,x,k,r\right)\right] dF_k dx,$$

and C comes from (5.10). The expectation is with respect to the distribution k of the relay and $z(\omega,x,k,d)$ is given by:

$$z(\omega,x,k,d) = \frac{1}{1+\omega\|x-d\|^{-\alpha}+\omega\|k-d\|^{-\alpha}}. \quad (6.14)$$

For $z(\omega,x,k,r)$ a similar expression holds interchanging d with r. The complexity of these expressions is due mainly to the interferences (6.5) and (6.6), and it precludes closed-form computations. For this reason we introduce the following far-field approximation for the path loss of the interfering clusters: the users within a cluster see the interference from other clusters as a point source of interference, meaning that:

$$\|x_i - r\| \approx \|x_i + k_i - r\| \approx \|x_i + \tau k_i - r\|, \quad (6.15)$$

$$\|x_i - d\| \approx \|x_i + k_i - d\| \approx \|x_i + \tau k_i - d\|. \quad (6.16)$$

The parameter τ allows to establish the far field approximation using any point between each source and its relay. As we shall see the results obtained are the same independently of its value. With this assumption a single path loss will appear in the interferences, so (6.5) and (6.6) can be simplified as:

$$\tilde{I}_r = \sum_{x\in\Phi_s} \frac{|h_{x_i r}|^2 + \varepsilon_{x_i}\left(|h_{k_i r}|^2 + 2\Re\left\{h_{x_i r}h^*_{k_i r}\rho\right\}\right)}{\|x_i+\tau k_i - r\|^\alpha}, \quad (6.17)$$

$$\tilde{I}_d = \sum_{x\in\Phi_s} \frac{|h_{x_i d}|^2 + \varepsilon_{x_i}\left(|h_{k_i d}|^2 + 2\Re\left\{h_{x_i d}h^*_{k_i d}\rho\right\}\right)}{\|x_i+\tau k_i - d\|^\alpha}. \quad (6.18)$$

This approximation will be very good in the high reliability regime because the independent fading coefficients are conserved and the large scale effect of path loss is still taken into

account. With these new interference expressions, we upper bound the OP of the network by introducing the union bound on the outage events of DF:

$$P_{\text{out,mix}}(R) \leq \mathbb{P}(\varepsilon_0 = 0)\mathbb{P}\left(\mathcal{A}_{DT}(R)|\varepsilon_0 = 0\right) + \mathbb{P}(\varepsilon_0 = 1) \times \mathbb{E}_r\left[\mathbb{P}\left(\mathcal{A}_{DF}(R)|r, \varepsilon_0 = 1\right) + \mathbb{P}\left(\mathcal{B}_{DF}(R)|r, \varepsilon_0 = 1\right)|\varepsilon_0 = 1\right]. \quad (6.19)$$

This will be a good approximation when the relay is not too far away from the source, since in that case the event $\mathcal{B}_{DF}(R)$ will be dominant and $\mathcal{A}_{DF}(R)$ will have a relatively small probability of occurrence. This will act as a good approximation to the OP of the network, and can be solved in closed form. In order to find the value of p_r which is minimized we shall resort to a big O expansion of this expression:

Theorem 6.2. *The OP upper bound (6.19) for this network can be evaluated as:*

$$P_{out,mix}(R) \leq (1-p_r)\left[1 - e^{-\lambda_s \Delta(p_r) D^2}\right] + p_r\left\{2 - \mathbb{E}_r\left[e^{-\lambda_s \Delta(p_r)||r||^2}\right] - e^{-\lambda_s \Delta(p_r) D^2} \left[1 + \lambda_s \Delta(p_r) D^2 \left(1 + \frac{2-\alpha}{\alpha D}\mathbb{E}_r[||r-d||]\right) + O\left((\lambda_s \Delta(p_r) D^2)^2\right)\right]\right\}, \quad (6.20)$$

as $\lambda_s \Delta(p_r) D^2 \triangleq \lambda_s \delta \left(1 + \frac{2p_r}{\alpha}\right) D^2 \to 0$, *with:*

$$\mathbb{E}_r\left[e^{-\lambda_s \Delta(p_r)||r||^2}\right] = \frac{\phi_0 \lambda_{in}}{\phi_0 \lambda_{in} + 2\lambda_s \Delta(p_r)}. \quad (6.21)$$

When $\phi_0 = 2\pi$ *we have a close form expression for the expectation:*

$$\mathbb{E}_r\left[||r-d||\right] = \sigma_{in} Q_{2,0}(D/\sigma_{in}, 0), \quad (6.22)$$

where $Q_{2,0}$ *is the* $(2,0)$ *Nuttall Q-function [74]:*

$$Q_{2,0}(s,0) = \sqrt{\frac{\pi}{8}} e^{-\frac{s^2}{4}} \left(\left(s^2+2\right) I_0\left(\frac{s^2}{4}\right) + s^2 I_1\left(\frac{s^2}{4}\right)\right),$$

where I_0 *and* I_1 *are the modified Bessel functions of the first kind of orders* 0 *and* 1. *In the general case we have to find the expectation numerically or we may use the following upper bound:*

$$\mathbb{E}_r\left[||r-d||\right] \leq D(1 + s\gamma(s, \phi_0)) \quad (6.23)$$

with $s = (\lambda_{in} \phi_0 D^2)^{-1/2}$ *and*

$$\gamma(s, \phi_0) = \sqrt{\frac{\pi}{2}}\left\{1 + \left[\frac{8(1-\cos(\phi_0/4))}{\phi_0} - 2\right] erf\left(\frac{1}{\sqrt{2}s}\right)\right\},$$

with $erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ *the standard error function.*

Proof. See Appendix C.1. □

6.4 Optimal relay activation probability

In the previous section we established an upper bound on the OP of the network choosing the relay as the nearest neighbor in a cone, as a function of the relay activation probability p_r and the cone aperture ϕ_0. For a given network set-up (R, α, λ_s, σ_{in}) different values of p_r and ϕ_0 will yield different values of the OP: increasing p_r will introduce additional interference in the network, while decreasing the cone aperture ϕ_0 will increase the average source-relay distance. If there is a high density of potential relays the cone aperture can be used to balance the average source-relay and source-destination distances to optimize the performance of the network. For this reason we should find the optimal values of p_r and ϕ_0, those which result in the smallest OP for each setup. In this section we study the optimal value of p_r in terms of the OP and determine the gains that can be achieved in terms of OP by optimizing this parameter.

Optimization of the relay activation probability using standard methods is very involved due to the non-linear nature of the expression of the OP. It would be expected that an optimal relay activation probability p_r would exist, which would optimally balance the effect of the added interference and the gains of activating additional relays.

Theorem 6.3. *Neglecting the term $O\left((\lambda_s \Delta(p_r) D^2)^2\right)$ in (6.20), for each network set-up (α, d, ϕ_0, λ_s, R) such that $\lambda_s \delta D^2 < \frac{3-\sqrt{5}}{2}$ there is an interval of σ_{in}:*

$$0 \leq \sigma_{in} \leq \sigma_c, \tag{6.24}$$

such that the OP upper bound is a concave function of p_r.

Proof. See Appendix C.2. □

Lemma 6.4. *Given a concave function $h(x)$ in a bounded and closed interval $[x_1, x_2]$, its minimum is attained at x_1 or x_2.*

Proof. See Theorem 32.1 in [75]. □

Since the upper bound will be a good approximation to the OP, using Lemma 6.4 together with Theorem 6.3 we conclude that the best OP performance for any cluster in the network can be attained when all ($p_r = 1$) or none ($p_r = 0$) of the sources decide to use their associated relays. In one case all the clusters will be using DF and in the other case all of them will be using DT. This is a somewhat surprising result in the sense that in terms of the OP the best performance can be obtained either by full cooperation or by not cooperating at all. There is no "optimal" density of used relays in the network or optimal mixed behavior in the sense that some clusters would enjoy the advantages of cooperation while others use DT, in order

to balance the generated interference. This interval clearly depends on the network set-up parameters; however, working in the realistic high reliability regime we can obtain a simple approximation of this condition that depends only on basic network parameters:

Corollary 6.5. *In the high reliability regime, we inner bound the concavity interval of Theorem 6.3 by finding the smallest positive solution to the equation:*

$$\frac{4\pi\alpha\sigma_c^2}{\phi_0 D^2} + (\alpha - 2)\,\frac{\mathbb{E}_r[||r-d||]}{D} - \alpha = 0. \tag{6.25}$$

Notice that $\mathbb{E}_r[||r-d||]$ *also depends on* σ_c *so the equation cannot be solved in closed form. By using (6.23) to upper bound the expectation, the following sufficient condition for a concave OP concave is obtained:*

$$\sigma_{in} \leq D\sqrt{\frac{\phi_0}{2\pi}}\left\{\left[\frac{1}{\alpha} + \varphi_c(\phi_0,\alpha)^2\right]^{1/2} - \varphi_c(\phi_0,\alpha)\right\}, \tag{6.26}$$

where:

$$\varphi_c(\phi_0,\alpha) = \frac{1}{4}\left(1 - \frac{2}{\alpha}\right)\gamma(1/\sqrt{2},\phi_0). \tag{6.27}$$

Proof. See Appendix C.3. □

So far we have established that there is a regime in which either $p_r = 1$ or $p_r = 0$ are the values that minimize the OP for a given network set-up (α, d, ϕ_0, λ_s, R), and in Corollary 6.5 we have determined conditions to find that interval. Now we wish to establish conditions under which we should activate all the relays, that is, when $p_r = 1$ will be the optimal choice:

Theorem 6.6 (Optimality region of $p_r = 1$)**.** *Neglecting the term* $O\left((\lambda_s\Delta(p_r)D^2)^2\right)$ *in (6.20), for each network set-up (*α*,* d*,* ϕ_0*,* λ_s*,* R*) such that* $\lambda_s\delta D^2 < \frac{3-\sqrt{5}}{2}$ *there is an interval of* σ_{in}*:*

$$0 \leq \sigma_{in} \leq \sigma_t \tag{6.28}$$

such that the OP upper bound is minimized by activating all the relays.

For the high reliability regime, an approximation for σ_t *is obtained by finding the smallest positive solution of the equation:*

$$\left[1 + \frac{2}{\alpha}\right]\left[\frac{4\pi\sigma_t^2}{\phi_0 D^2} + \left(1 - \frac{2}{\alpha}\right)\frac{\mathbb{E}_r[||r-d||]}{D}\right] = 1. \tag{6.29}$$

Notice that the expectation also depends on σ_t *so the equation is coupled. By using (6.23) to upper bound the expectation, the following sufficient condition for* $p_r = 1$ *to be optimal is obtained:*

$$\sigma_{in} \leq D\sqrt{\frac{\phi_0}{2\pi}}\left\{\left[\frac{2}{\alpha(\alpha+2)} + \varphi_t(\phi_0,\alpha)^2\right]^{1/2} - \varphi_t(\phi_0,\alpha)\right\}, \tag{6.30}$$

where:

$$\varphi_t(\phi_0,\alpha) = \frac{1}{4}\left(1 - \frac{2}{\alpha}\right)\gamma(1/2,\phi_0). \tag{6.31}$$

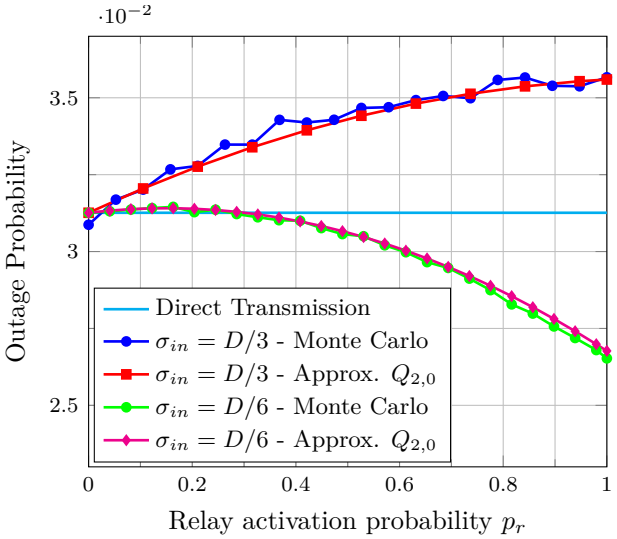

Figure 6.3: Outage probability $P_{\text{out,mix}}(R, p_r)$ as a function of p_r for values of σ_{in} showing optimality of $p_r = 0$ or $p_r = 1$. $d = (10,\ 0)$, $\lambda_s = 10^{-4}$, $R = 0.5$, $\alpha = 4$. Monte Carlo simulations are obtained by averaging 8×10^6 realizations of the PPP using (6.5) and (6.6). Approximations come from using (6.22) in (6.20).

Proof. See Appendix C.4. □

Using the previous theorems, we are able to state a relay activation scheme that optimizes the OP in a network operating in the high reliability regime: for a given value of ϕ_0 if σ_{in} is less than the solution of (6.29) then all the relays should be on. Otherwise, the relays should be turned off and DT should be employed. A computationally simpler alternative for turning the relays on would be using condition (6.30) instead. The value of ϕ_0 could additionally be chosen within this scheme to minimize the OP. Notice that $\sigma_t \equiv \sigma_t(\phi_0)$ is a function of ϕ_0. If for a network set-up (α, d, λ_s, R) there is a value of ϕ_0 such that $p_r = 1$ is optimal, i.e. $\sigma_{in} \leq \sigma_t(\phi_0)$ holds, then there will be a range of values of ϕ_0 for which this condition will hold. We should therefore choose the value of ϕ_0 for which $\sigma_{in} < \sigma_t(\phi_0)$ holds and the OP is minimized. On the other hand, if there is no value of ϕ_0 such that $\sigma_{in} < \sigma_t$ we have that $p_r = 0$ is optimal and hence, DT should be employed.

As we shall observe in the section of numerical results, there will be scenarios in which setting $\phi_0 = 2\pi$ will yield approximately the same performance as optimizing the value of ϕ_0 in terms of the OP according to the previous observation. This means that, in practical scenarios, this optimization may not always be of importance and the value of σ_t can be obtained by setting $\phi_0 = 2\pi$ in (6.29).

Finally, we want to compare the OP that can be achieved with the scheme defined in Theorem 6.6 with the one obtained using only DT. For each value of ϕ_0, if $p_r = 1$ minimizes the OP then the scheme will exhibit gains with respect to DT, while if $p_r = 0$ is the optimum, the scheme reverts to DT, and no gains will be seen. The following theorem finds the approximate reduction of the OP of the scheme with respect to DT:

Theorem 6.7. *In the high reliability regime, the relative decrease in OP of the activation scheme obtained by using (6.29) is:*

$$\frac{P_{out,mix}}{P_{out,DT}} \approx \begin{cases} \left(1+\frac{2}{\alpha}\right)\left(\frac{4\pi\sigma_{in}^2}{\phi_0 D^2}+\left(1-\frac{2}{\alpha}\right)\frac{\mathbb{E}_r[||r-d||]}{D}\right) & \sigma_{in} \leq \sigma_t, \\ 1 & \textit{otherwise.} \end{cases} \tag{6.32}$$

In addition σ_t can be lower bounded by (6.30).

Proof. See Appendix C.5. □

As we mentioned before, both σ_t and the actual reduction in OP are a function of ϕ_0. If ϕ_0 can be optimized, then for each network setup (α, d, λ_s, R) we have to determine (if they exist) the values of ϕ_0 such that $\sigma_{in} \leq \sigma_t$ (which ensure a gain with respect to DT) and from those values, the one that minimizes the OP.

6.5 Numerical results

In this section we present some simulations to study the behavior of the expressions we have introduced previously. In Fig. 6.3 the OP with respect to p_r is plotted for two different values of σ_{in}, one in which $p_r = 1$ is optimal and another one for which $p_r = 0$ is the optimal point, when the relay is selected as the nearest neighbor on the whole plane ($\phi_0 = 2\pi$). The theoretical expressions come from the upper bound (6.20) using (6.22), and they are compared with Monte Carlo simulations obtained by averaging 8×10^6 realizations of the PPP using the true interferences (6.5) and (6.6), taking $d = (10,\ 0)$, $\lambda_s = 10^{-4}$, $R = 0.5$ bit/use and $\alpha = 4$. We see that the approximations derived with the simplified interferences (6.17) and (6.18) are in excellent agreement with the actual OP derived with the more complex interferences.

In Fig. 6.4 we plot the optimal cone aperture ϕ_0 as a function of σ_{in}/D for different values of the path loss exponent α and for $d = (10, 0)$. To do this, we numerically find the value of ϕ_0 that maximizes the OP gain of the mixed protocol with respect to DT in (6.32) for each value of σ_{in}/D. It is interesting to note that as the path loss exponent decreases the optimal cone aperture becomes $\phi_0 = 2\pi$ for a large range of values of σ_{in}/D. Only when the network of potential relays is very dense (small σ_{in}/D) a value of $\phi_0 < 2\pi$ should

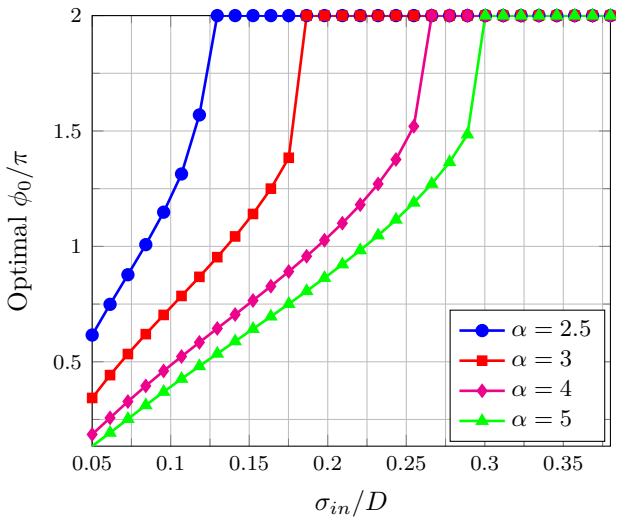

Figure 6.4: Optimal cone aperture ϕ_0 as a function of σ_{in}/D obtained using (6.32) for different values of α. $\lambda_s = 10^{-4}$. $d = (10, 0)$.

be chosen. This is because when the exponent diminishes both the source-relay and the interference paths become stronger, but the effect of the increased interference is dominant. Thus the diminished exponent creates an effect equivalent to increasing the average source-relay distance. The value of ϕ_0 must therefore become larger in order to decrease the average source-relay distance and compensate for this effect.

In Fig. 6.5 we study the maximum rate attainable for the on/off relaying strategy relative to the same rate of DT in percentage for a desired OP value of 0.03. The maximum rates are obtained by using (6.20). For the plots with $\phi_0 = 2\pi$ the rates are obtained by using (6.22) while in the other case the expectations are computed numerically. For the plots with optimized cone aperture we use the values of ϕ_0 from Fig. 6.4, taking $\lambda_s = 10^{-4}$ and $d = (10, 0)$. The on/off condition (which predicts when the rate of the mixed scheme reaches that of DT) is obtained by solving (6.29). We have also plotted as vertical lines the simpler on/off condition (6.30) which is in excellent agreement with the other one. We observe that optimizing the cone aperture can be helpful when the path loss exponent or the density of potential relays are large. In addition, as the path loss exponent decreases we can achieve a lower maximum rate with DT for a given outage constraint; this implies that the benefits of a reduced exponent within the cluster are outmatched by the simultaneous increase in interference due also to the reduced exponent. The plot also shows that although the maximum rate for DT is smaller, the relative gains of the mixed scheme become larger. This means that the maximum achievable rate decreases slower for the mixed scheme than

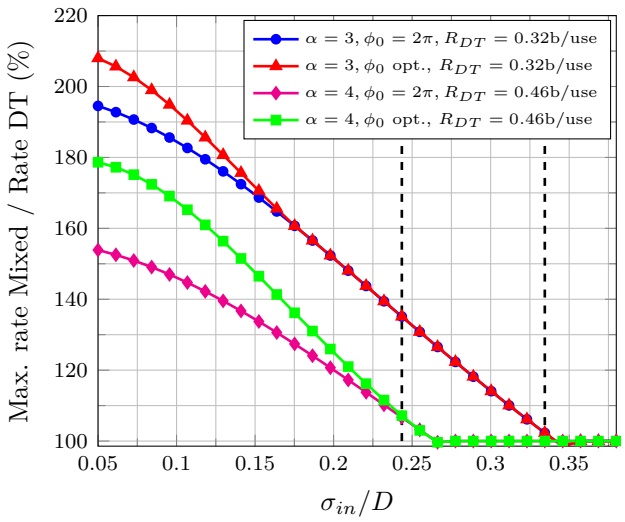

Figure 6.5: Maximum rate attainable for the on/off relaying strategy relative to the same rate of DT for a given OP constraint. The mixed scheme rates are obtained by using (6.20) and the on/off condition from solving (6.29). We also plot the on/off condition (6.30). The optimal aperture angles come from Fig. 6.4. $\lambda_s = 10^{-4}$. $d = (10, 0)$.

for DT as the path loss exponent decreases, which suggests that the increased interference is less damaging for the mixed scheme than for DT.

In Fig. 6.6 we plot the relative gain in OP with respect to DT as a function of σ_{in}/D using $\phi_0 = 2\pi$ and the optimal cone apertures from Fig. 6.4. The OP gains are obtained from theorem 6.7. We also plotted as vertical lines the simpler condition (6.30) which is in excellent agreement with the other one. Finally, in Fig. 6.8 we compare the performance of the proposed on/off strategy against with two other simple relay activation schemes: one in which the relay is activated if the source-relay channel exceeds a threshold and another one in which a threshold on the relay-destination channel is used. Both schemes make use of the available CSI. In the first case, the relay can determine if the threshold is exceeded, and in the second one, the destination, who has CSI on the relay-destination link, can send a bit (at negligible cost) indicating if the relay should transmit or not. In both cases, the path loss and the corresponding fading coefficients are considered. The OP curves for these schemes are determined through Monte Carlo simulations of the point process and for each point the value of the threshold is numerically optimized to obtain the smallest OP possible. These curves are compared to the OP from the upper bound (6.20) and the on/off strategy. For these simulations we use $\lambda_s = 10^{-4}$, $R = 0.5$ b/use, $d = (10, 0)$, $\alpha = 4$. We observe that although these schemes employ available CSI which is not taken into account by the

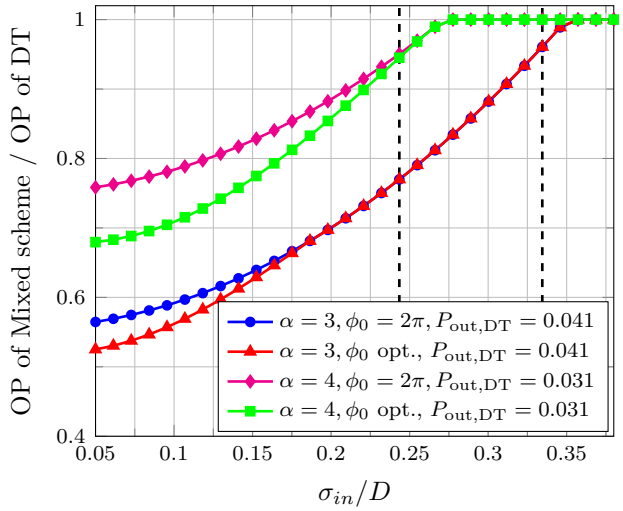

Figure 6.6: Relative improvement in OP with respect to DT for the on/off scheme as a function of σ_{in}/D as predicted by Theorem 6.7. We have also plotted as vertical lines the on/off condition (6.30). $d = (10, 0)$.

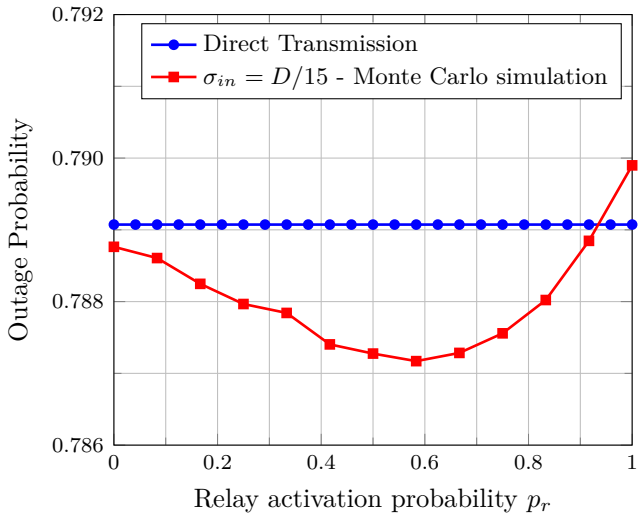

Figure 6.7: Outage probability $P_{\text{out,mix}}(R, p_r)$ as a function of p_r showing that the OP is not always a concave function of p_r. $d = (70,\ 0)$, $\sigma_{in} = D/15$, $\lambda_s = 10^{-4}$, $R = 0.5$, $\alpha = 4$. Monte Carlo simulations are obtained by averaging 4×10^6 realizations of the PPP using the true interferences.

independent activation schemes, the performance is similar between the three strategies.

Finally, it is interesting to mention that under certain conditions the OP is not a concave

function of p_r, that is, the OP is minimized by choosing a value of p_r which is different from $p_r = 0$ or $p_r = 1$. However, in such scenarios the network is well outside the high reliability regime and typical operating conditions. To show this, we provide Fig. 6.7 which is a Monte Carlo simulation of the PPP (without any simplification of the interference functions) for $\alpha = 4$ and $R = 0.5$b/use (the same parameters used in Fig. 6.3). We can see that the optimal value of p_r is not either $p_r = 0$ or $p_r = 1$. However, the network is operating at an OP of the order of 0.79 which is well out of the operation level of usual wireless systems.

6.6 Summary and final remarks

In this work we analyzed the performance of a large wireless network under a mixed cooperative randomized scheme which employs either DF or DT, and obtained the optimal relay activation strategy for this network. When DF is used, the relays are chosen as the nearest neighbor within a cone, with its axis towards the destination. This is a natural assumption since DF is known to be near optimal when the relay is not too far from the source. At the same time, the effect of the path loss on the relay-destination link, which is very detrimental to the performance of the scheme, is reduced. The choice between DT and DF is done by the corresponding relay associated with each source via a randomized decision with probability p_r and without taking into account any additional knowledge the relays might have. This simple procedure, which is mathematically tractable, can be thought as a MAC layer at the relays (in a similar fashion as the popular ALOHA protocol), with the objective of limiting the interference generation in the network. On the other hand it could also model a situation in which the relays are unavailable due to conditions out of control of the source or the relay itself, such as, for example, a depleted battery. With this simple model, a balance between cooperation and interference generation can be established in the network. Surprisingly, for typical operating conditions, the optimal values of p_r are 0 or 1, revealing a binary behavior: all nodes in the network should use their relay or none at all. Following this conclusion, a relay activation strategy was introduced to achieve the optimal behavior. Even when cooperation is beneficial to all, the performance improvements may not be as large as in the typical fading relay channel with Gaussian noise. The reason for this comes from the fact that, in addition to fading, we have averaged over all possible node configurations, including many cases in which interference is very damaging. It is interesting to mention that the model introduced and several results, such as Theorem 6.2, can be used to study other relay selection and activation algorithms based on position, such as choosing the relay as the nearest or farthest neighbor on a finite cone, and with minor modifications extend them to other cases involving additional CSI. Other protocols assuming higher degrees of CSI may yield better gains, but

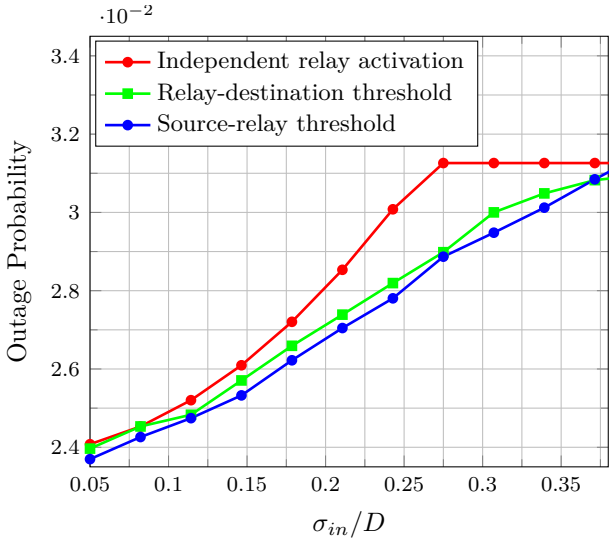

Figure 6.8: Comparison between the OP attainable through independent relay activation and through the use of a threshold on the source-relay or source-destination channel. $\lambda_s = 10^{-4}$, $R = 0.5$ b/use, $d = (10, 0)$, $\alpha = 4$. The independent relay curve comes from (6.20) and the rest from Monte Carlo simulations of the PPP. Thresholds are optimized numerically for best performance.

this may not be a realistic assumption in this context. A potential improvement could be obtained using more sophisticated cooperative transmission schemes which could take into account the impairments generated by the nearby interferers [76] (which introduce by far the most harmful interference). Basically, this could consist on decoding the messages sent by strong nearby interferers first, then subtracting them from the received signal and finally attempting to decode the desired message. In such situation, besides the intrinsic benefits of cooperation, the smart use of the aggregate interference introduced in part by the cooperating nodes, could ameliorate its harmful effect on the overall network. Another improvement could be obtained by using more advanced MAC schemes for the relays and the sources, such as CSMA, to avoid simultaneous nearby transmissions. In this context a metric such as the transmission capacity [42] may be more appropriate for the analysis. Finally, the study of other cooperative schemes as AF and CF deserves full consideration. All these issues, as well as the effect of using several potential relays instead of only one, constitute important and interesting future work directions.

Chapter 7

On Fundamental Trade-offs of Device-to-Device Communications in Large Wireless Networks

Summary. This chapter studies the gains, in terms of served requests, attainable through out-of-band D2D video exchanges in large cellular networks. A stochastic framework, in which users are clustered to exchange videos, is introduced, considering several aspects of this problem: the video-caching policy, user matching for exchanges, aspects regarding scheduling and transmissions. A family of *admissible protocols* is introduced: each protocol is composed of a *clustering strategy* induced by any hard-core point process, and any suitable *in-cluster communication strategy*, which dictates the dynamics inside the clusters. Two metrics, quantifying the "local" and "global" fraction of video requests served through D2D are defined, and relevant trade-off regions involving these metrics, as well as quality-of-service constraints are identified. A simple communication strategy is proposed and analyzed, to obtain inner bounds to the trade-off regions, and draw conclusions on the performance attainable through D2D. To this end, an analysis of the time-varying interference that the nodes experience, and tight approximations of its Laplace transform are derived.

7.1 Introduction

7.1.1 Motivation and related work

Cellular D2D communications, in which two or more mobile users establish a direct link without going through the BS, have emerged as a viable alternative to partially cope with the increasing demands in data traffic and connectivity that cellular networks will face in the future. Generally speaking, D2D communications are opportunistic, one-hop, short range transmissions in which the BS can be used for communication and acts as a last case fall-back alternative [19]. This allows –among other things– for a higher spatial frequency reuse, energy efficiency, coverage extension, and a reduced load on the backhaul.

The scope of D2D communications is very wide, from machine-to-machine (M2M), gaming

and relaying, to content distribution or even public safety networks [18, 19]. Among these, an important application is that of video content distribution. This is because, in the next few years, traffic from wireless and mobile devices will exceed traffic from wired devices, and this will be largely related to an increase in *video on demand* and Internet-to-TV downloads, which can have a multiplier effect on traffic [1]. The asynchronous nature of VoD requests implies that, in many cases, multicasting strategies for video transmissions cannot be employed, even though a small library of videos may be accessed by many users [20]. An approach that has been recently proposed to mitigate this consists in including small distributed caching stations with a limited backhaul that can locally serve video video demands [77, 78]. Another approach [20] is to take advantage of the low-cost largely-unused storage space that is available in many wireless devices to store and exchange videos locally. In this way, either the users could cache videos they have already seen or the service providers could select the approriate videos to cache and distribute them during moments of low traffic in the network. In this chapter we focus on this distributed caching approach, which, in principle, would not require dedicated storage units. Our main goal is to study the potential benefits that could be obtained through a distributed user-caching strategy, by considering the fraction of mean video requests that could be served through D2D, without requiring the BS to transmit them. This may yield some insight on the impact of D2D in terms of video availability and backhaul load, which may have implications both economically and in terms of quality of service. To this end, we introduce a simple framework for analysis, based on a stochastic geometry model. In this framework, users are assumed to be grouped into clusters where D2D video exchanges take place. We attempt to consider both the problem of establishing matches between nearby requesting and caching users, and the problem of scheduling and transmission, involving slow fading, path loss, and interference between nodes. We focus specifically on *out-of-band* D2D, which utilizes bands other than the cellular ones, thus increasing the frequency reuse and mitigating interference in the cellular band. We also discuss the trade-offs between the fraction of requests which are served “locally” (per cluster) and “globally" (in an arbitrary region) in the network. It is important to mention that the idea is to catch a small library of videos which are on high demand during a certain time period in the network, and not to catch all the videos that a user may request, which would be practically impossible.

Some related works which focus on local video caching and D2D exchanges include [20, 78, 79, 80]. Broadly speaking, these works consider finite area networks in which a fixed number of users are distributed either uniformly or on a regular grid. The model for transmission failure is generally the distance-based *protocol model* introduced in [62]. They also consider *out-of-band* D2D but focus mainly on the optimal asymptotic scaling laws of the networks and on design-parameter optimization. For example, in [20, 80] a one-hop network in which

users are clustered and cache videos is studied with this model, and a throughput-outage trade-offs is characterized for various regimes, in terms of scaling laws as the number of nodes and the number of videos in the library grow to infinity. In [79] the authors find an optimal collaboration distance to balance interference, and analyze the scaling behavior of the benefits of D2D. On the other hand, our approach considers an infinite-area constant-density model in which transmissions are impaired by path-loss and fading.

Other works regarding the performance of D2D communications involving stochastic geometry models are [81, 82, 83]. In general, these works are not focused specifically on video distribution in D2D networks which we attempt to analyze, but on general traffic and general aspects of D2D communications. For this reason they do not consider the problem of user matching, user request statistics and caching policies, which becomes central in the video distribution problem. Actually, in [81] the authors study the optimal downlink spectrum partition between D2D and BS transmissions through a time-frequency hopping scheme for scheduling D2D. In [82], the authors analyze a D2D *in-band overlaid* cellular network model and find expressions for important performance metrics, such as the outage probability.

7.1.2 Main contributions

The main goal of this chapter is to study the number of requests that could be served by D2D instead of asking the BS for transmission. To this end, we propose a stochastic geometry framework, with the following characteristics:

- Requesting users (destinations) and cooperative users (with cached videos) are distributed in space as a Poisson process on the plane. Transmissions are affected by path loss, slow fading and interference.

- Users are grouped in disjoint clusters where D2D exchanges take place. A family of *admissible protocols* is introduced: each protocol is composed of a *clustering strategy* induced by any hard-core point process [2], and any suitable *in-cluster communication strategy*, which dictates the communication schemes of users inside the clusters.

In this setting, we define two metrics of interest, which characterize the performance of D2D in terms of served requests:

- A *global* metric that measures the ratio between the spatial density of served requests and the total density of requests. This gives the global fraction of the video requests which could be served through D2D exchanges without using the downlink of the cellular network.

- A *local* metric that measures the ratio between the average number of served requests and total requests in a cluster. This can be interpreted as an indication of what gains could be expected in a localized region in space, in which certain level of service is expected.

Although these metrics pertain the three aspects of the problem mentioned earlier, introducing a link-quality constraint is reasonable to model the quality of service constraints, such as large data rate, which may be required in video distribution. Otherwise, a large fraction of requests could be satisfied through low-rate transmissions, which may be not be reasonable in this setup. For this reason, we introduce three trade-off regions pertaining these metrics:

- The *global metric-average rate* trade-off region, which pertains the fraction of requests than can be served considering an average rate requirement over the cluster.
- The *global metric-average rate and cluster density* trade-offs region, which refers to the local fraction of requests than can be served considering that an average rate and a minimal cluster density are required.
- The *global-local* trade-off regions, which pertain the balance between the global and the local density of served requests which are attainable simultaneously.

Determining these regions implies characterizing the optimal communication scheme among the family of admissible protocols mentioned before. Since optimal communication schemes remain unknown for each trade-off region, we analyze a simple in-cluster communication strategy, which can be paired with any clustering strategy to obtain a protocol. This will give an inner bound to the trade-off regions. In this strategy, users which request videos and those with cached videos are paired and a one-hop transmission takes place. Interference within clusters is avoided by precluding simultaneous transmissions through a time-division multiple access (TDMA) scheme which shares the time resource between transmitters. It is shown that the TDMA scheme implies that a user will experience a time-varying interference during a slot. An analysis which takes this into account is performed, in order to determine the rates that a user can achieve. This analysis, which is not usually considered in a stochastic geometry setup, may be of interest for scenarios other than D2D. We then evaluate the global and local metrics for all the protocols obtained by pairing this in-cluster communication strategy with any clustering strategy. In this way, we obtain a set of inner bounds to the optimal trade-off regions, which give an indication of the possible gains through D2D. Finally, we numerically evaluate one of this inner bounds, by considering the clustering strategy induced by a type II Matérn hard-core point process [2] and Rayleigh fading. Then, by developing tight approximations to the Laplace transform of the interference anywhere in a cluster, we

derive the trade-off regions and draw conclusions on the performance attainable through D2D video exchanges.

The chapter is organized as follows. In Section 7.2, we introduce the network model, the family of admissible protocols and the main metrics and trade-offs under study. In Section 7.3, we introduce and analyze the simple communication strategy and derive the approximations on the Laplace transform of the interference. In Section 7.4, we present some relevant plots and comments, while in Section 7.5 we discuss our findings. Finally, proofs are relegated to Appendix D.

7.2 System model and admissible protocols

In Section 7.2.1 we introduce a simple model of how users are distributed and cache videos. Also, we introduce a family of clustering strategies, which group the users into clusters, in which D2D exchanges take place. This spatial model is fully characterized through an independently marked point process. In Section 7.2.2 we define a family of in-cluster communication strategies, which define how users may interact inside the cluster to exchange videos. In Section 7.2.3 we introduce a set of metrics which allow us to study the performance attainable through the set of clustered D2D strategies.

7.2.1 Clustering strategies and spatial model

We consider an infinite planar network in which:

- Users who request videos are distributed according to an homogeneous Poisson process Φ_r, of intensity λ_r.
- Users who cache videos are distributed according to an homogeneous Poisson process Φ_u, of intensity λ_u[1].
- Users attempt to exchange videos through D2D to reduce the load on the downlink of the cellular network. This is done outside the downlink frequency band so there is no interference between cellular and D2D communications.

[1]Notice that we can consider these two processes as originating from a single Poisson process of intensity $\lambda_u + \lambda_r$, where then a user decides to become a requesting user with probability $\lambda_r/(\lambda_u + \lambda_r)$, independent of everything else, and the rest are caching users. The separation between requesting and caching users is done to simplify the model. In reality, requesting users would also cache videos, which would increase the likelihood of finding videos and the number of served requests (provided an efficient medium access and transmission strategy are considered).

- Each user in Φ_r requests a video which is selected independently according to a discrete distribution $p_V(v)$, where $1 \leq v \leq L$ and L is the library size. In numerical results, we assume that the videos are sorted according to their popularity which implies that $p_V(v)$ is the probability of requesting the v-th most popular video. This distribution is commonly [20, 78, 79, 80] assumed to be a Zipf distribution of parameter $0 < \gamma < 1$:

$$p_V(v) = \frac{v^{-\gamma}}{\sum_{i=1}^{L} i^{-\gamma}}, \quad v \in \{1, \ldots, L\}. \tag{7.1}$$

- Each user of Φ_u has M (fixed) videos cached, which, for simplicity, are selected independently according to a discrete distribution $p_A(a)$, $1 \leq a \leq L$. For numerical results we assume that $p_A \equiv p_V$; this could be motivated by assuming that users cache the files they have watched.

- Transmissions are subject to both slow fading and path loss. The power received at y by a unit-power transmission from x is $|h_{xy}|^2 l(x, y)$ where $l(x, y) = \|x - y\|^{-\alpha}$ ($\alpha > 2$) is the usual path loss function and $|h_{xy}|^2$ is the power gain of fading with unit mean. In our analysis we focus on the interference generated between the nodes using D2D, and focus on the signal-to-interference ratio (SIR). Independent heterogeneous background interference or noise could be added in a straightforward manner.

The model could also consider that users cache parts of videos instead of whole videos, but it would only increase the set of possible messages and the size of the cache, without introducing substantial differences in the actual model. Also, the assumption that some users cache videos and others request videos is considered mainly for simplicity (to avoid the events that users have the video they request), but for the analysis we make it could be relaxed.

To exchange videos, users of Φ_r and Φ_u are grouped into disjoint clusters which, for simplicity, are approximated as discs of radius R_c. The users of Φ_r which are not clustered will need to ask for the videos directly to the BS. On the other hand, the users of Φ_r which are clustered can search inside their cluster for a user from Φ_u who has the video and request a transmission through D2D. We assume that this grouping lasts for at least the duration of a communications block (it may change from block to block). The assumption that the disk-shaped clusters do not intersect implies there is a minimum distance of at least $2R_c$ between the centers and hence, modeling the cluster centers as a hard-core point process [2], which guarantees this clearance, is a natural assumption. This leads to the following definition.

Definition 7.2.1 (Process of clustered users)**.** *Given a cluster radius $R_c > 0$, a process of clustered users Φ_c is constructed from Φ_u and Φ_r as follows:*

$$\Phi_c = \bigcup_{x \in \Phi_p} \mathcal{B}(x, R_c) \cap (\Phi_r \cup \Phi_u), \tag{7.2}$$

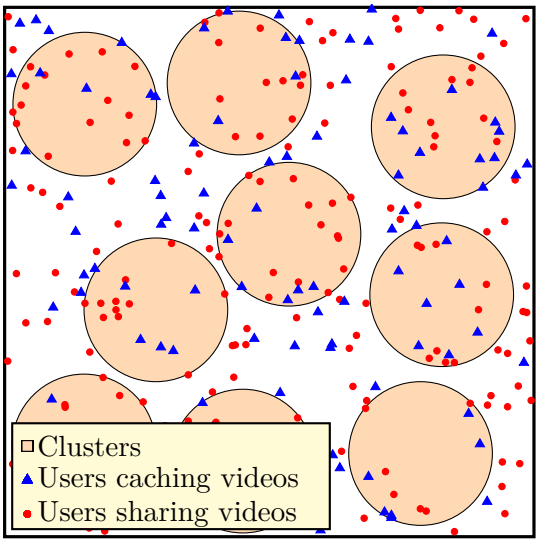

Figure 7.1: Representation of the network with clusters. The users in the clusters, form the cluster process Φ_c according to (7.2), in which D2D takes place.

where $\Phi_p = \{x_i\}$ *is a stationary* parent *hard-core process of intensity* $\lambda_p > 0$ *and clearance distance* $\delta \geq 2R_c$.

Notice that any hard-core stationary point process with clearance $\delta \geq 2R_c$ will generate a cluster process. In Fig. 7.1 we can see a representation of the network.

Remark 7.2.1. *Under the previous hypotheses, the same model can be constructed if we first deploy the discs of radius* R_c*, then create inside each disc independent Poisson processes of the same intensities as* Φ_u *and* Φ_r*, and then create PPPs outside the space occupied by the clusters.*

Each cluster will have a family of associated users which cache or request videos, which will be points of the original processes Φ_u and Φ_r. Each of these users will be represented by some information, mainly, their positions, the video(s) they cache or request, and fading coefficients towards other users in the network. We can represent this information of the users in each cluster as a vector of random variables which is associated to the point of the cluster center, and, due to the assumptions we have stated earlier in the section, these marks will be independent among clusters. This means we can represent the network of clusters as a stationary independently marked point process (Section 4.3):

$$\tilde{\Phi} = \{(x_i, \mathbf{m}_{x_i})\}, \tag{7.3}$$

where $\Phi_p = \{x_i\}$ is the process of cluster centers, given by a hard-core process of intensity λ_p, and $\mathbf{m}_x$ is a mark vector, containing all the random variables of the cluster at x, which

are:

- $N_{x,u}$: the number of users which cache videos in the cluster centered at x. They are Poisson random variables of mean $\lambda_u \pi R_c^2$. $\mathbf{S}_x = (S_{x,1}, \ldots, S_{x,N_{x,u}})$ is the vector of positions of these users relative to x, which, conditioned on $N_{x,u}$ are independent uniform random variables on the cluster (Remark 7.2.1).

- $N_{x,r}$: the number of users requesting videos within the cluster centered at x. They are Poisson random variables of mean $\lambda_r \pi R_c^2$. $\mathbf{D}_x = (D_{x,1}, \ldots, D_{x,N_{x,r}})$ is the vector of positions of these users relative to x, which, conditioned on $N_{x,r}$ are independent uniform random variables on the cluster (Remark 7.2.1).

- $\tilde{\mathbf{A}}_x = \{\mathbf{A}_{x,1}, \ldots, \mathbf{A}_{x,N_{x,u}}\}$ are the videos which are stored in the users, such that $S_{x,i}$ stores $\mathbf{A}_{x,i} = (A_{x,i,1}, \ldots, A_{x,i,M})$. They are selected independently according to p_A as indicated before. $\mathbf{V}_x = \{V_{x,1}, \ldots, V_{x,N_{x,r}}\}$ are the requested videos such that $D_{x,i}$ requests $V_{x,i}$. They are selected independently according to p_V, as mentioned earlier.

- $\mathbf{H}_x$ is a family of independent power fading coefficients between the users inside the cluster and towards users in other clusters.

When clear from context the subscript x in $\mathbf{m}_x$ is omitted. The dependence between the variables in the mark vector is characterized by their joint distribution $F_{\mathbf{m}_x}$. The caching and requesting users are distributed as independent Poisson processes in the clusters (see Remark 7.2.1), and considering the independent requests and caching strategy, this distribution can be factored as:

$$
\begin{aligned}
F_{\mathbf{m}_x} &= F_{N_{x,u},N_{x,r},\mathbf{S}_x,\mathbf{D}_x,\tilde{\mathbf{A}}_x,\mathbf{V}_x,\mathbf{H}_x} && (7.4)\\
&= F_{\mathbf{S}_x,\mathbf{D}_x,\tilde{\mathbf{A}}_x,\mathbf{V}_x,\mathbf{H}_x|N_{x,u},N_{x,r}} F_{N_{x,u},N_{x,r}} && (7.5)\\
&= F_{\mathbf{H}_x|\mathbf{S}_x,\mathbf{D}_x,N_{x,u},N_{x,r}} F_{\mathbf{S}_x|N_{x,u}} F_{\tilde{\mathbf{A}}_x|N_{x,u}} F_{\mathbf{D}_x|N_{x,r}} F_{\mathbf{V}_x|N_{x,r}} F_{N_{x,u}} F_{N_{x,r}} && (7.6)\\
&= F_{N_{x,u}} F_{N_{x,r}} \left(\prod_{i=1}^{n_r} F_{V_{x,i}|N_{x,r}=n_r} F_{D_{x,i}|N_{x,r}=n_r} \right) \left[\prod_{i=1}^{n_u} F_{S_{x,i}|N_{x,u}=n_u} \left(\prod_{j=1}^{M} F_{A_{x,i,j}|N_{x,u}=n_u} \right) \right] \\
&\qquad \times F_{\mathbf{H}|\mathbf{S}_x,\mathbf{D}_x,N_{x,u},N_{x,r}}. && (7.7)
\end{aligned}
$$

For shortness, unless mandatory like in the last step, we have not included the point where the distributions are evaluated. For example, we haven written $F_{\mathbf{D}_x|N_{x,r}} \equiv F_{\mathbf{D}_x|N_{x,r}=n_r}(\mathbf{d}_x)$.

It is important to mention that there are several hard core processes, with different characteristics, which can be used as a parent process Φ_p. They offer different degrees of regularity and fractions of network coverage. Two important examples are Matérn's hard-core models [2, 84], which are obtained from a PPP Φ of intensity λ, in which certain points are deleted

through a position-dependent thinning. In type I processes, any two points which are separated by less than the clearance distance δ are deleted. In type II processes, a uniform random variable is drawn for each point of Φ and, for each pair of points which are separated by less than δ, only the one with the smallest uniform random variable remains. Another example is the translated-grid model [26, 7], which gives a regular square grid. It is obtained by using two independent uniform random variables in $[0, \delta)$, U_1 and U_2, and by considering the grid formed by the pairs $(m\delta + U_1, n\delta + U_2)$, for all integers m, n. With these rules, parent processes which guarantee a minimum clearance of $\delta \geq 2R_c$ are obtained, and their densities are:

$$\lambda_p = \lambda e^{-\lambda\pi\delta^2} \qquad \text{Matérn Type I,} \tag{7.8}$$

$$\lambda_p = \frac{(1 - e^{-\lambda\pi\delta^2})}{\pi\delta^2} \qquad \text{Matérn Type II,} \tag{7.9}$$

$$\lambda_p = \frac{1}{\delta^2} \qquad \text{Translated grid.} \tag{7.10}$$

Other processes may be constructed via usual operations on point processes, like thinning [2].

7.2.2 In-cluster interaction and admissible protocols

Given a realization of $\tilde{\Phi}$, in each cluster, a memoryless network is defined, where users who cache videos are sources, which have a subset of all possible messages, and requesting users are receivers, requesting one of the possible messages each. To conserve the network symmetry, keep a simple structure and not require long-range coordination, we assume that:

- Transmitters may have different degrees of CSI pertaining only to their own cluster, that is, some knowledge about the cluster, which is contained in its mark vector, $\mathbf{m}_x$.
- Clusters are uncoordinated, interference between them is treated as noise and there is no interaction between them to reduce interference. This implies the behavior of the users in a cluster is defined only by its mark vector, $\mathbf{m}_x$. Then, to define the behavior of the clusters, we may introduce a function $\zeta(\mathbf{m})$ which assigns a multiuser coding scheme to every realization of $\mathbf{m}$ (the same function for all clusters). This means that any two clusters in the network having the same realization of their mark vectors, will use the same coding scheme, given by ζ. Also, since the mark vectors $\{\mathbf{m}_{x_i}\}_i$ are independent and identically distributed, the same would be true for the coding scheme selection random variables $\{\zeta(\mathbf{m}_{x_i})\}_i$.
- Transmissions in the network take place in a block, in which all the clusters attempt to serve some or all of the requests inside at the same time. A transmission of rate R

is required for transmitting any of the videos, and a unit average-power constraint is imposed on each user.

With the above assumptions, it is enough to focus on a single cluster to describe the behavior of any cluster in the network. Let us suppose that conditioned on $\tilde{\Phi}$ we pick a cluster $(x, \mathbf{m}_x)$ which contains n_u users with videos and n_r requests for which the video is available in the cluster. The cluster uses the coding scheme $\zeta(\mathbf{m}_x)$ and this defines an achievable rate region $\mathcal{R}(\mathbf{m}_x, \tilde{\Phi}) \subset \mathbb{R}_+^{n_r}$ of the rates that are achievable from the n_u transmit nodes to the n_r receiver nodes. This is not the region in which all the users decode successfully, but an extended region, which is the union of the regions in which at least one of the users can decode. Then, we may define a vector of information rates $\mathbf{r} = (R, \ldots, R) \in \mathbb{R}_+^{n_r}$, in which the i-th component is the rate attempted through the coding scheme towards receiver i. If $\mathbf{r} \in \mathcal{R}(\mathbf{m}_x, \tilde{\Phi})$ then all the users will decode successfully, while if this does not happen, then a subset of the users will be able to decode. Due to the symmetry, the asymptotic error probability of each user (after averaging over $\tilde{\Phi}$) will be the same.

Definition 7.2.2 (In-cluster communication strategy)**.** *An in-cluster communication strategy is a set of admissible coding schemes for the involved network (one for each realization of* $\mathbf{m}$*) in which a symmetric rate* R *is attempted to all users.*

Definition 7.2.3 (Served request)**.** *A request from the* i*-th user in a cluster centered at* x *is said to be served if the following three things happen:*

- *The video is available in the cluster, that is, there is a match for this user, an event which can be written as:*
$$\mathcal{M}_{x,i} = \bigcup_{j=1}^{N_{x,u}} \{V_{x,i} \in \mathbf{A}_{x,j}\}. \tag{7.11}$$
- *The transmission is scheduled during the block, that is, the user with the match is scheduled to receive a transmission from one or more of the users with the video.*
- *The* i*-th transmission rate* R *belongs to the achievable rate-region* $\mathcal{R}(\mathbf{m}_x)$ *induced by the strategy.*

The probability of having a match is the same for all the users in the network, so we can write $\mathbb{P}(\mathcal{M}_{x,i}) \equiv p_{\mathcal{M}}$. However, within a cluster, these events are correlated for different users, because they search for the videos in the same distributed cache.

We now define the family of admissible protocols:

Definition 7.2.4 (Admissible protocol)**.** *An admissible protocol is any pair* $(\Phi_p, \mathcal{F})$*, where* Φ_p *is an admissible parent process, which defines a clustered network* $\tilde{\Phi}$ *and* $\mathcal{F}$ *is an in-cluster communication strategy as in Def. 7.2.2.*

7.2.3 Performance metrics and trade-offs

For every admissible protocol $(\Phi_p, \mathcal{F})$ and given a compact set $K \subset \mathbb{R}^2$ we define the random variable $N_s(K, \Phi_p, \mathcal{F}, R)$ as the number of served requests in K during a block, which we can write as:

$$N_s(K, \Phi_p, \mathcal{F}, R) \triangleq \sum_{x \in \Phi_p} \sum_{i=1}^{N_{x,r}} \mathbb{1}_{\mathcal{K}_{x,i}} \mathbb{1}_{\mathcal{S}_{x,i}}, \tag{7.12}$$

where:

$$\mathcal{S}_{x,i} = \{\text{Req. of user } i \text{ in } (x, \mathbf{m}_x) \text{ is served}\}, \tag{7.13}$$

$$\mathcal{K}_{x,i} = \{x + D_{x,i} \in K\}. \tag{7.14}$$

In addition, we define $N_{sc}(\Phi_p, \mathcal{F}, R, x)$ as the number of served requests in a cluster centered at x, that is:

$$N_{sc}(\Phi_p, \mathcal{F}, R, x) \triangleq \sum_{i=1}^{N_{x,r}} \mathbb{1}_{\mathcal{S}_{x,i}}. \tag{7.15}$$

Lemma 7.1. *Given a compact set $K \subset \mathbb{R}^2$, the average number of served requests in K is:*

$$\mathbb{E}\left[N_s(K, \Phi_p, \mathcal{F}, R)\right] = \lambda_p |K| \mathbb{E}^0[N_{sc}(\Phi_p, \mathcal{F}, R, 0)], \tag{7.16}$$

where $|K|$ denotes the area of K. $\mathbb{E}^0$ is the expectation with respect to the Palm distribution of the point process with a cluster at the origin, interpreted as a conditional distribution which averages over the realizations of the point process with a cluster at the origin. The term $\mathbb{E}^0[N_{sc}(\Phi_p, \mathcal{F}, R, 0)]$ represents the average number of users served within any cluster in the network.

Proof. It is an application of the Campbell-Mecke formula (4.23). Details are included in Appendix D.1. □

We next define the main metrics under study.

Definition 7.2.5 (Local metric)**.** *The ratio of mean served requests per cluster is:*

$$T_L(\mathcal{F}, R) = \frac{\mathbb{E}^0[N_{sc}(\Phi_p, \mathcal{F}, R, 0)]}{\mathbb{E}[N_r]}, \tag{7.17}$$

where $\mathbb{E}[N_r] = \lambda_r \pi R_c^2$ is the average number of requests within any cluster of the network. This ratio indicates how many requests are served on average in any cluster of the network, relative to the average number of requests per cluster.

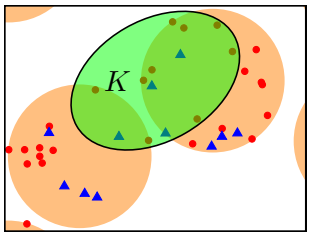

Figure 7.2: For a compact set K, the global metric measures the ratio of served requests in K (clustered) and total requests in K (clustered or not).

Linking this metric with (7.16) we define a global metric:

Definition 7.2.6 (Global metric)**.** *Chosen a compact set K, the ratio of mean served requests is:*

$$T_G(\mathcal{F}, R) = \frac{\mathbb{E}\left[N_s(K, \Phi_p, \mathcal{F}, R)\right]}{\mathbb{E}[N_r(K)]} \tag{7.18}$$

$$= \lambda_p |\mathcal{B}(0, R_c)| T_L(\mathcal{F}, R), \tag{7.19}$$

where $\mathbb{E}[N_r(K)] = \lambda_r |K|$ is the average number of requests in the set K. Due to the stationarity of the process, this metric does not depend on K and it can be interpreted as the spatial density of served requests, normalized by the density of requests λ_r.

The local metric refers to the level of service in a cluster located somewhere, that is, it is a metric which conditions a cluster being located somewhere. On the other hand, the global metric considers an arbitrary set K which will contain both clustered and unclustered requests, and it will consider as unserved requests those which fall outside a cluster. Therefore, if the local metric is close to one with a vanishing density of clusters, the global metric will be zero, because the global, i.e., considering unclustered requests, fraction of requests will be zero (see Fig. 7.2). The metrics are determined by factors such as the caching policy and video request statistics, the attempted rate R, and the parameters of the parent process, such as the cluster radius, and the in-cluster communication strategy. There is trade-off behavior in these metrics. The local metric always benefits from a reduction in the cluster density because as $\lambda_p \to 0$ the interference decreases on average. If the cluster radius R_c is fixed and λ_p is diminished, the local metric will benefit, but if the density becomes too small, the global metric will eventually have to decrease.

Definition 7.2.7 (Average rate)**.** *Given an admissible protocol $(\Phi_p, \mathcal{F})$, a cluster centered at*

x has an average rate:

$$\bar{R}(\mathcal{F}, R) \triangleq R\ \mathbb{E}^x \left[\sum_{i=1}^{N_{x,r}} \frac{\mathbb{1}_{\mathcal{S}_{x,i}}}{\sum_{j=1}^{N_{x,r}} \mathbb{1}_{\mathcal{M}_{x,j} \cap \mathcal{P}_{x,j}}} \right]. \tag{7.20}$$

where $\mathcal{P}_{x,i}$ is the event that a transmission to user i is scheduled in the time block. This average rate is the same for all x due to the stationarity of the process.

We consider the metrics in Defs. 7.2.5 and 7.2.6, together with an average-rate constraint, to model requirements in terms of delay and link-reliability, which may be required in the context of video distribution. For this reason, we introduce the following trade-off regions:

Definition 7.2.8 (Trade-off regions)**.** *We consider the following regions:*

- Global-metric trade-off region: *a pair (r, t) of average rate and global metric is said to be achievable if there exists an admissible protocol $(\Phi_p^\dagger, \mathcal{F}^\dagger)$ with rate R and density λ_p satisfying:*

$$\begin{cases} T_G(\mathcal{F}^\dagger, R) \geq t, \\ \bar{R}(\mathcal{F}^\dagger, R) \geq r. \end{cases} \tag{7.21}$$

  The set of all achievable pairs (r, t) is the global-metric trade-off region.

- Local-metric trade-off region: *a tuple (r, t, λ_l) of average rate, local metric and parent density is said to be achievable if there exists an admissible protocol $(\Phi_p^\dagger, \mathcal{F}^\dagger)$ with rate R and density λ_p satisfying:*

$$\begin{cases} T_L(\mathcal{F}^\dagger, R) \geq t, \\ \bar{R}(\mathcal{F}^\dagger, R) \geq r, \\ \lambda_p(R_c, \delta) \geq \lambda_l. \end{cases} \tag{7.22}$$

  The set of all achievable tuples (r, t, λ_l) is the local-metric trade-off region.

- Local-global trade-off region: *given an attempted rate R, a pair (t_g, t_c) of global and local metrics is said to be achievable if there exists an admissible protocol $(\Phi_p^\dagger, \mathcal{F}^\dagger)$ with rate R and density λ_p satisfying:*

$$\begin{cases} T_G(\mathcal{F}^\dagger, R) \geq t_g, \\ T_L(\mathcal{F}^\dagger, R) \geq t_c. \end{cases} \tag{7.23}$$

  The set of all achievable pairs (t_g, t_c) is the local-global trade-off region.

The first region refers to the maximum fraction of users which receive videos successfully globally, subject to an average rate constraint. This metric is limited by the fraction of the

users of the network that are clustered, because unclustered users cannot exchange videos. A point process $\Phi_p^\dagger$ should be chosen such that the network is almost fully covered by clusters, and a strategy $\mathcal{F}^\dagger$ such that all the requests in a cluster can be served ($T_L \approx 1$), while fulfilling the rate constraint. The second region refers to the fraction of the users inside a cluster that receive videos successfully. In this case, both a rate constraint and a certain density of clusters are required. Otherwise, we could set $\lambda_p \approx 0$ and achieve a large level of service at the typical cluster, but there would be almost no other cluster in the network. Since this region is defined by what happens in a cluster, we could have $T_L \approx 1$, for any clustering process. The third region refers to maximizing one the metrics, with a constraint on the other one, balancing the global and local benefits of D2D.

We cannot find the optimal protocol which is close to optimality in terms of each trade-off region. However, analyzing a specific protocol will yield inner regions and thus yield insights on the performance attainable through D2D.

The following lemma provides a straightforward bound for the local and global metrics.

Lemma 7.2. *Given a protocol $(\Phi_p, \mathcal{F})$, we have*

$$T_L(\mathcal{F}, R) \leq p_{\mathcal{M}},$$

with equality as $R \to 0$. This implies for the global metric that

$$T_G(\mathcal{F}, R) \leq \lambda_p p_{\mathcal{M}} |\mathcal{B}(0, R_c)|.$$

In addition, under the caching scheme described, the probability of a match is:

$$p_{\mathcal{M}} = 1 - \mathbb{E}_V^0 \left[e^{-\lambda_u \pi R_c^2 [1-(1-p_A(V))^M]} \right]. \tag{7.24}$$

Proof. See Appendix D.2. □

This lemma is quite intuitive: if no constraint is imposed on the transmission quality, the local metric T_L is limited by the probability of finding a match in a cluster $p_{\mathcal{M}}$, and the global metric T_G is limited by this probability and the fraction of clustered users, which is $\lambda_p |\mathcal{B}(0, R_c)|$. If we introduce additional constraints on the average rate or the density of the parent process, the bounds will still be valid, so they will give outer bounds on the regions achievable if we fix the parent process.

7.3 Analysis of an in-cluster communication strategy

7.3.1 Strategy definition

In this section we define and analyze an in-cluster communication strategy which, paired with any parent process, forms an admissible D2D protocol via Def. 7.2.4. We focus on the cluster

at the origin $(0, \mathbf{m}_0)$, omitting the subscript 0 in all its random variables ($N_{0,u} \equiv N_u$, etc.). To consider a simple strategy we assume that:

- To reduce the interference inside the clusters, a TDMA scheme is employed to divide the transmission block into equal sized slots, in which only one request is served, through a point-to-point unit-power transmission using Gaussian signaling.
- At most $n_{m,\max}$ slots are defined in each cluster, regardless of the number of matches. This does not have any practical implications, since we can choose this such that $F_{N_m}(n_{m,\max}) \approx 1$, i.e., fraction of clusters with dropped requests is negligible[2].

To fully occupy the block with this TDMA scheme, each cluster would split the block into as many matches as it has, and clusters with the same number of matches would have the same number of slots. This assignment is reasonable in terms of theoretical performance, but leads to an interference model which is not tractable, mainly because the transmissions between clusters with a different number of matches are unsynchronized. Figure 7.3, which focuses on a cluster with one slot, is provided to help understand this. When a slot is over in another cluster, its transmitter is replaced by a new one, which changes a part of the interference. Meanwhile, the other clusters will generate the same interference as before. This results in a time-correlated interference, whose statistics are very involved to model, specially considering the large range of number of slots in a very large network. Also, the distribution of the interference would not be the same in all slots of the cluster. This is inherent to any wireless system using a similar TDMA scheme.

To overcome this, some degree of regularity is required. Splitting the block in $n_{m,\max}$ slots would solve the problem but lead to an inefficient use of resources, since the block would be mostly unoccupied in all clusters. For this, we propose a strategy in which the block always has a power of two number of slots. This allows a different number of slots between clusters, improves the use of resources, and considers interference changes during a slot.

Definition 7.3.1 (Coordination strategy)**.** *In a cluster with N_m matches:*

- *The block is split into $W(N_m, \varepsilon)$ slots:*

$$W(N_m, \varepsilon) = \begin{cases} W_L(N_m) & \textit{when } \frac{N_m - W_L(N_m)}{W_H(N_m) - W_L(N_m)} < \varepsilon, \\ W_H(N_m) & \textit{otherwise,} \end{cases} \tag{7.25}$$

[2]This requirement is considered for mathematical reasons. In any cluster, the number of slots/matches is always finite, but, since there is an infinite number of clusters, the maximum number of slots over all clusters is unbounded, which is in conflict with the finite (yet long) length of the block which is required to achieve reliable communications.

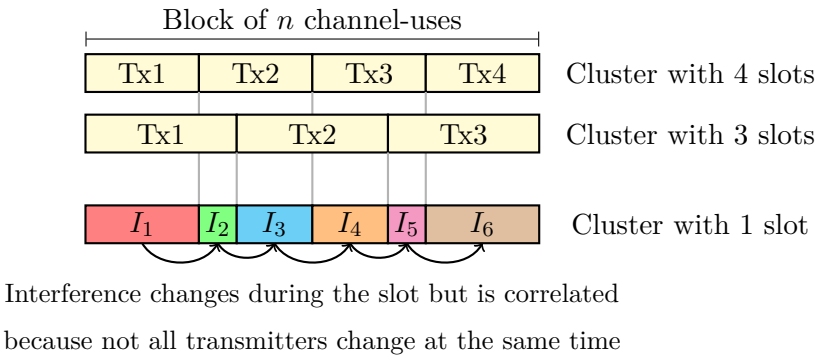

Figure 7.3: If each cluster employs a TDMA scheme, dividing the transmission block in a number of slots equal to the number of matches, interference becomes time-correlated and non-stationary. This is because transmissions among cluster would be unsynchronized.

where $0 < \varepsilon \leq 1$ *is a design parameter, and:*

$$W_H(N_m) = 2^{\lceil \log_2(N_m) \rceil}, \;\; W_L(N_m) = 2^{\lfloor \log_2(N_m) \rfloor}, \tag{7.26}$$

the powers of two closest to N_m *from above and below, respectively. The random variables* $W(N_m, \varepsilon) \equiv W_x(N_{x,m}, \varepsilon)$*, defined for each cluster, are independent because they are a function of the* $\{N_{x_i,m}\}_i$*, which also are independent.*

- *For each request with a match, a caching user is selected at random from the set of users who have the video:*

$$\tilde{\mathcal{A}}(V_i, \tilde{\mathbf{A}}) \triangleq \{j : V_i \in \mathbf{A}_j, 1 \leq j \leq N_u\}, \tag{7.27}$$

as a candidate to serve the request.

- *If there are* $W = W_H(N_m)$ *slots, transmissions are scheduled by selecting* N_m *out of the* $W_H(N_m)$*, and generating a random permutation of the transmissions in these slots. Otherwise,* $N_m - W_L(N_m)$ *requests are dropped at random, and the rest are assigned by generating a random permutation of the slots.*

Figure 7.4 is presented to illustrate how the protocol works: given a number of matches N_m, the closest powers of two above and below N_m are found, $W_H(N_m)$ and $W_L(N_m)$, respectively. The parameter ε defines a threshold and balances the fraction of time in which the channel is occupied per cluster, with the fraction of the requests that are served. When $\varepsilon = 0$, $W_H(N_m) \geq N_m$ slots are used, meaning all the requests are served, possibly leaving some slots empty. When $\varepsilon = 1$ some requests may be dropped because $W_L(N_m) \leq N_m$ slots are used, but all the slots will be occupied. Also, by dropping requests, a higher success

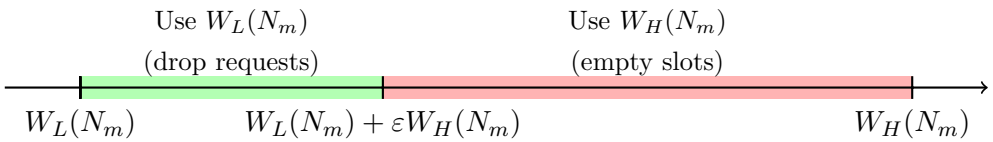

Figure 7.4: Representation of the slotting scheme proposed. Given a number of matches N_m, ε defines a threshold; if the number of matches is below it, requests are dropped but all slot are occupied. Otherwise, all requests are scheduled but there may be empty slots.

probability per user may be achieved because each transmission gets more channel uses[3]. Since only one transmitter is chosen at random from $\tilde{\mathcal{A}}(V_i, \tilde{\mathbf{A}})$ to serve a request, for each user $1 \leq i \leq N_r$ in a cluster, we define an random variable C_i indicating which user transmits:

$$C_i|_{V_i,\tilde{\mathbf{A}}} \sim U(\mathcal{A}(V_i, \tilde{\mathbf{A}})) \quad \text{if } \tilde{\mathcal{A}}(V_i, \tilde{\mathbf{A}}) \neq \emptyset, \tag{7.28}$$

where $U(\cdot)$ denotes a uniform random variable over the set, and $C_i = 0$ if $\tilde{\mathcal{A}}(V_i, \tilde{\mathbf{A}}) = \emptyset$.

7.3.2 Interference characterization and achievable rates

We now analyze the interference generated by our scheme, in order to define the achievable rates of the users. To do this, we follow the same approach as in the previous chapters by conditioning on the point process $\tilde{\Phi}$, but also on the realization of the scheduling strategy in Def. 7.3.1 and focus on a cluster centered at the origin. Having conditioned on the realization of the strategy, there will be a number of slots defined in this cluster $N_{0,m} \geq 0$; we will focus on one of them and refer to it as the *slot under study*.

Assume that the cluster at the origin has n_1 slots, n_1 being a power of two, as indicated in Def. 7.3.1. Fig. 7.5 shows how the interference behaves during the slot under study:

- The other clusters in the network which have at most n_1 slots generate a constant interference power because only one transmitter is active during the slot under study.
- Clusters with more than n_1 slots, say $2^k n_1$, will generate 2^k interference powers during the slot under study.

The maximum number of slots in a cluster is:

$$\Delta \triangleq W_H(n_{m,\max}), \tag{7.29}$$

[3]We haven't considered the possibility that $N_m > n_{m,\max}$. We could do this, but it would further complicate the exposition without adding any substantial modifications, because both the probability of this event in the typical cluster, and the fraction of the clusters in this condition are negligible.

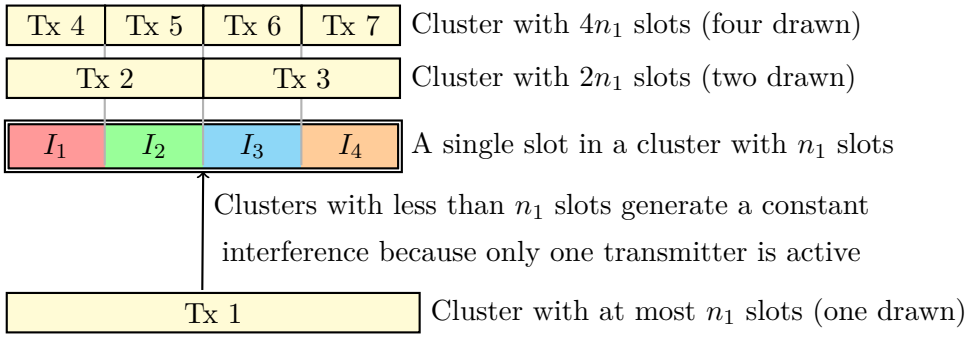

Figure 7.5: The double line box is a single slot in a cluster with n_1 slots. Clusters with more than n_1 slots cause time variations in the interference, because different transmitters are active in each slot. Clusters with at most n_1 slots generate a constant interference, because only one transmitter is active.

which means that every cluster in the network will see $W_H(n_{m,\max})$ different interference power during the time-block. Since the slot under study will occupy a fraction of $1/n_1$ of the total block length, it will see Δ/n_1 different interference powers. Under the Gaussian signaling assumption, a communication during the slot will be a point-to-point transmission in which the interference powers change Δ/n_1 times. If $\{\tilde{I}_1(y), \ldots, \tilde{I}_{\Delta/n_1}(y)\}$ are these powers, and we consider a long block-length, a transmission from x to y could achieve a rate R:

$$R < \frac{1}{n_1}\left(\frac{n_1}{\Delta}\sum_{i=1}^{\Delta/n_1} \mathsf{C}\left(\mathrm{SIR}_i(x,y)\right)\right) = \frac{1}{\Delta}\sum_{i=1}^{\Delta/n_1} \mathsf{C}\left(\mathrm{SIR}_i(x,y)\right), \tag{7.30}$$

where

$$\mathrm{SIR}_i(x,y) = \frac{|h_{xy}|^2 l(x,y)}{\tilde{I}_i(y)}.$$

The outer fraction $1/n_1$ represents that the slot will take up the fraction $1/n_1$ of the block, while the inner terms represent that during the slot there will be n_1/Δ different interference powers. Notice that SIR changes are caused by the interference, while the source-destination channel is the same.

The complexity of (7.30) precludes finding the probability of a failed transmission, i.e. the probability that (7.30) does not happen, so we develop a lower bound on the achievable rate. Since we can classify the clusters according to the number of slots that they have, $\{1, 2, 4, ..., \Delta\}$, and the clusters with at most n_1 slots will generate a constant interference during the slot under study, we can define the aggregate interference generated by all these clusters $I_b(n_1, y)$ as:

$$I_b(n_1, y) = \sum_{j=0}^{\log n_1} I_j(y), \tag{7.31}$$

where $I_j(y)$ is the aggregate interference at y from the clusters with 2^j slots during the slot under study. Clusters with more than n_1 slots, say $2^k n_1$, give 2^k interference values; we index these 2^k values using a binary expansion, as $I_k(u_1, \ldots, u_k, y)$, with $(u_1, \ldots, u_k) \in \{0,1\}^k$, for any $1 \leq k \leq \log_2(\Delta/n_1)$. This indexing is only used because it is convenient for the proof of Lemma 7.3. We do not need to specify and order in which these interferences appear during the slot under study; we are only interested in the time average of these values:

$$\bar{I}_k(n_1, y) = \frac{1}{2^k} \sum_{(u_1,\ldots,u_k)\in\{0,1\}^k} I_k(u_1, \ldots, u_k, y). \tag{7.32}$$

That is, $\bar{I}_k(n_1, y)$ is the time-average of the interference coming from clusters with $2^k n_1$ slots, seen at y during the slot under study. With these definitions we introduce the following:

Lemma 7.3 (Achievable rate)**.** *The rate (7.30) of any slot in a cluster with a total of n_1 slots can be lower bounded by:*

$$R_a(n_1, x, y) = \frac{1}{n_1} \mathsf{C}\left(\frac{|h_{xy}|^2 l(x,y)}{I_b(n_1, y) + \sum_{i=1}^{\log_2(\Delta/n1)} \bar{I}_k(n_1, y)} \right), \tag{7.33}$$

where I_b comes from (7.31) and $\bar{I}_k(n_1, y)$ is given by (7.32).

Proof. See Appendix D.3. □

This means that any rate $R < R_a$ is achievable, where R_a is a time-sharing, point-to-point rate with a constant interference, made up of the original interferences added together, with components that change over the slot being time-averaged. We now write the expression of the interference in (7.33) as seen from the slot under study. If we focus on an interfering cluster with $2^k n_1$ slots, we see that the same transmitter can be active in more than one slot, or some slots may be empty, so we cannot assume that there will be 2^k different or non-null interference powers (coming from each cluster). Since we have time-averaged the interference, we are only interested in how many times each user transmits, and not in the transmission order. By assuming that all the slots are occupied, which gives a worst case, the interference in (7.33) at a point d of the cluster at the origin with n_1 slots, during a slot, is:

$$I(d, n_1) = \sum_{x\in\Phi_p\setminus\{0\}} \left\{ \mathbb{1}\{W_x \leq n_1\} |h_{xd}|^2 l(x + S_x, d) + \right.$$
$$\left. \sum_{i=\log(n_1)+1}^{\log\Delta} \mathbb{1}\{W_x = 2^i\} \frac{n_1}{2^i} \sum_{j=1}^{2^i/n_1} B_{x,j,i} |h_{x+S_{x,j},d}|^2 l(x + S_{x,j}, d) \right\}. \tag{7.34}$$

The terms with $\mathbb{1}\{W_x \leq n_1\}$ represent the interference from users in clusters with at most n_1 slots. The terms with $\mathbb{1}\{W_x = 2^i\}$ represent the interference from a cluster with 2^i slots.

The random variables $B_{x,j,i} \geq 0$ indicate how many times each transmitters is active during slot. Since we take that all the slots are occupied, these variables have to satisfy:

$$\sum_{j=1}^{2^i/n_1} B_{x,j,i} = 2^i/n_1. \tag{7.35}$$

If in a cluster $B_{x,j,i} = 1$, for all j, then every slot that took place during the slot under study is used by a different transmitter; if only one satisfies $B_{x,j,i} = 2^i/n_1$ and the rest are zero, all the slots are used by the same transmitter.

Remark 7.3.1. *Actually, there is an abuse of notation in (7.34). The terms* $|h_{x+S_{x,j},d}|^2 l(x+S_{x,j},d)$ *do not use the same indexing over* j *as in the definition of* $\tilde{\Phi}$*. In fact, for example, for a cluster with* $2^i > n_1$ *slots we cannot guarantee that there will be* $2^i/n_1$ *different users with videos stored, i.e. that the sum over* j *is well defined. What we do is obtain an upper bound to the interference seen at* d *during the slot under study. To do this, we take each cluster, see which transmitters will be active during the slot under study, and count how many times each will transmit. Then, we may add more fictitious transmitters to have* $2^i/n_1$ *transmitters, and assign them slots such that (7.35) is met. We do not need to consider which of the users actually transmits in each cluster, because we are focused on a single slot of the cluster at the origin, and because, conditioned on the cluster centers, the fading and user positions random variables are all independent. The distribution of the variables* $\{B_{x,j,i}\}$ *will not be required, because we will show that assuming that all transmitters are different, that is, taking* $B_{x,j,i} = 1$ *for all* j*, is a worst case in terms of the interference.*

7.3.3 Performance metric analysis

We now evaluate the metrics from Def. 7.2.8 for the strategy given by Def 7.3.1, which for simplicity we denote by $\mathcal{F}^*$. Then for a compact set K and a protocol $(\Phi_p, \mathcal{F}^*)$, the average number of served requests is given by (7.12). The event of a served requests $\mathcal{S}_{x,i}$ can be written as: $\mathcal{S}_{x,i} = \mathcal{M}_{x,i} \cap \mathcal{T}_{x,i}$, where $\mathcal{M}_{x,i}$ indicates a match, and $\mathcal{T}_{x,i}$ means a transmission was scheduled and was successful ($R < R_a$, R_a from (7.33)).

Theorem 7.4 (Local and global metrics)**.** *Given a parent process* Φ_p*, as in Def. 7.2.1, the ratio of mean served requests per cluster, or local metric* T_L *(7.17), for* $\mathcal{F}^*$ *is given by:*

$$T_L(\mathcal{F}^*, R) = \frac{\mathbb{E}^0\left[\left(N_m \mathbb{1}_{\{W(N_m)=W_H(N_m)\}} + W_L(N_m)\mathbb{1}_{\{W(N_m)=W_L(N_m)\}}\right)\bar{F}_{|h|^2}\left(\frac{I(D,W(N_m))(2^{W(N_m)R}-1)}{l(S,D)}\right)\right]}{\mathbb{E}[N_r]}, \tag{7.36}$$

where $\bar{F}_{|h|^2}$ is the complementary distribution function of the fading coefficients. The spatial density of served requests or global metric (7.19) for the strategy $\mathcal{F}^*$ is therefore:

$$T_G(\mathcal{F}^*, R) = \lambda_p |\mathcal{B}(0, R_c)| T_L(\mathcal{F}^*, R). \tag{7.37}$$

The average rate achieved is:

$$\bar{R}(\mathcal{F}^*, R) = R\ \mathbb{E}^0\left[\bar{F}_{|h|^2}\left(\frac{I(D, W(N_m)(2^{W(N_m)R}-1)}{l(S,D)}\right)\right], \tag{7.38}$$

where $W(N_m)$ is given by (7.25). For Rayleigh fading:

$$\mathbb{E}^0\left[\bar{F}_{|h|^2}\left(\eta I(d)\right)\right] = \mathcal{L}^0_{I(d,n)}(\eta) \tag{7.39}$$

where $\mathcal{L}^0_{I(d,n)}(\eta) = \mathbb{E}^0\left[e^{-\eta I(d,n)}\right]$, is the Laplace transform [8] of $I(d,n)$ with respect to $\mathbb{P}^0$.

Proof. See Appendix D.4. □

The term $\bar{F}_{|h|^2}(\cdot)$ in (7.36) gives the probability of a successful transmission from S to D when the cluster has N_m matches and an interference $I(D, W(N_m))$. The number of requests are Poisson random variables with $\mathbb{E}[N_r] = \lambda_r \pi R_c^2$.

It is interesting to mention that achieving a high level of service within a cluster (local benefits) may imply that only a few clusters need be created, and hence, globally the effects of D2D may not be significant. From a local perspective, larger clusters imply that users are more likely to have matches, but also increases the average distance between users and reduces the time resource assigned to each transmission, increasing the chances of failed transmissions.

We now focus on the Rayleigh fading case, which requires the Laplace transform of the interference (7.34). The main issue in this case is the distribution of the variables $B_{x,j,i}$. Fortunately, the following lemma helps us avoid this:

Lemma 7.5. *The Laplace transform $\mathcal{L}^0_{I(z,n)}$ of (7.34) can be lower bounded by setting $B_{x,j,i} = 1$ for all x, j, k.*

Proof. See Appendix D.5. □

Notice that these results are valid for any process of cluster centers Φ_p. We now consider that Φ_p is a type II Matérn hard-core process and Rayleigh fading is present.

7.3.4 Approximations and bounds for Matérn type II processes

The main issue to evaluate the metrics for a type II process is finding the reduced Laplace transform of the interference (7.39). This process has been used mostly to model networks

using carrier-sense multiple access schemes [8, 67], and even the most simple transforms are not known in closed form. For this reason, we approximate the process by a more tractable one, following the approach in [85], in which the interference of a type II hard-core process, as seen from a node at the origin, is compared to that of a non-homogeneous PPP with a zero intensity in a ball around the origin. In our setting, this approach is equivalent to considering that the cluster centers, as seen from the cluster at the origin, to be distributed as a non-homogeneous Poisson process of intensity:

$$\lambda_p \mathbb{1}_{\{||x||>\delta\}},$$

where λ_p, given by (7.9), is the intensity of the original hard-core process. If we use Lemma 7.5 to bound the true interference (7.34), and approximate the Matérn process in this way, we have the following approximate interference:

$$I(d,n) \approx \hat{I}(d,n) = \sum_{x\in\Phi} \tilde{\psi}(x, \mathbf{m}_x, d, n)\mathbb{1}_{\{||x||>\delta\}}, \tag{7.40}$$

where $\tilde{\psi}(x, \mathbf{m}_x, d, n)$ is the function in the sum (7.34) with all the $B_{x,i,j} = 1$:

$$\tilde{\psi}(x, \mathbf{m}_x, d, n) = \mathbb{1}\{W_x \le n\}|h_{xd}|^2 l(x + S_x, d) + \sum_{i=\log n+1}^{\log \Delta} \mathbb{1}\{W_x = 2^i\}\frac{n}{2^i}\sum_{j=1}^{2^i/n} |h_{x,j,d}|^2 l(x + S_{x,j}, d). \tag{7.41}$$

The sum in (7.40) is over an homogeneous PPP of intensity λ_p and the non-homogeneity is given by $\mathbb{1}_{\{||x||>\delta\}}$. For this process we know that the reduced Laplace transform is the same as Laplace transform, so using Lemma 4.3 we would have:

$$\mathcal{L}_{\hat{I}(d,n)}(\eta) = \exp\left\{-\lambda_p \int_{||x||>\delta} \left(1 - \mathbb{E}\left[e^{-\eta\tilde{\psi}(x,\mathbf{m}_x,d,n)}\right]\right)dx\right\}, \tag{7.42}$$

where the inner expectation is over $\mathbf{m}_x$. Taking the expectation over W_x and over the independent unit-mean exponential fading random variables we have:

$$\mathbb{E}\left[e^{-\eta\tilde{\psi}(x,\mathbf{m}_x,d,n)}\right] = \mathbb{P}(W_x \le n)\mathbb{E}\left[\frac{1}{1 + l(x + S_x, d)\eta}\right] + \sum_{i=\log_2 n+1}^{\log_2 \Delta} \mathbb{P}(W_x = 2^i)\prod_{j=1}^{2^i/n} \mathbb{E}\left[\frac{1}{1 + l(x + S_{x,j}, d)\eta n/2^i}\right], \tag{7.43}$$

where the remaining expectations are over S_x, $\{S_{x,j}\}_j$, which cannot be computed in closed form. However, it is reasonable to introduce the far field approximation (used already in Chapter 6) that the interfering users are seen from the typical cluster as if located at the center

of their cluster. This is because the favorable and unfavorable positions of the interferers will be approximately canceled, because they are uniformly distributed around the center. With this approximation and using that:

$$\lceil 2^i/n \rceil = \begin{cases} 1 & \text{if } i \leq \log_2 n, \\ 2^i/n & \text{if } \log_2 n < i \leq \Delta, \end{cases} \tag{7.44}$$

(7.43) can be approximated as:

$$\mathbb{E}\left[e^{-\eta\tilde{\psi}(x,\mathbf{m}_x,d,n)}\right] \approx \sum_{i=0}^{\Delta} \frac{\mathbb{P}(W_x = 2^i)}{\left(1+\frac{l(x,d)\eta}{\lceil 2^i/n \rceil}\right)^{\lceil \frac{2^i}{n} \rceil}}. \tag{7.45}$$

Replacing this equation in (7.42), we have:

$$\mathcal{L}^0_{I(z,n)}(\eta) \approx \mathcal{L}_{\hat{I}(d,n)}(\eta) \approx \exp\left\{-\lambda_p \sum_{i=0}^{\Delta} \mathbb{P}(W = 2^i) \int\limits_{||x||>\delta} \left[1 - \frac{1}{\left(1+\frac{l(x,d)\eta}{\lceil 2^i/n \rceil}\right)^{\lceil \frac{2^i}{n} \rceil}}\right] dx\right\}. \tag{7.46}$$

The summation in (7.46) will have few terms ($\Delta = \log_2 n_{m,\max}$) and hence is straightforward to implement numerically.

7.4 Relevant plots and comments

In this section we study the inner bounds to the trade-off regions in Def. 7.2.8 which are achievable through the protocol $(\Phi_{HC}, \mathcal{F}^*)$ (Theorem 7.4) where Φ_{HC} is a type II hard-core process of intensity λ_p. Other processes of cluster centers, such as the ones mentioned in Section 7.2.1 may also be used and may provide a better performance. For all simulations, the requesting user density is $\lambda_r = 0.003$ users/area, the caching user density is $\lambda_u = 4\lambda_r$. We assume that $p_A \equiv p_V$, V follows a Zipf distribution of parameter $\gamma = 0.6$, and each caching user stores $M = 6$ videos. The path loss exponent is $\alpha = 4$. We take $\varepsilon = 0.05$ which means we favor leaving some slots unoccupied over dropping matches, which showed to be better. We use (7.46) to approximate the Laplace transform of the interference. Monte Carlo simulations are performed to average over S and D and to estimate the distribution of N_m.

In Fig. 7.6 we plot the inner bound of the global metric trade-off region (7.21) achievable by $(\Phi_{HC}, \mathcal{F}^*)$, changing the library size L. Omitting the dependence on $(\Phi_{HC}, \mathcal{F}^*)$, the region is:

$$\max_{R_c,R,\lambda,\delta} T_G(R, R_c, \delta, \lambda_p) \text{ subject to } \bar{R}(R, R_c, \lambda, \delta) \geq r. \tag{7.47}$$

Optimizing over all parameters R_c, R, λ, δ implies this is the best global fraction of served requests achievable by $(\Phi_{HC}, \mathcal{F}^*)$. Locally, increasing L requires, on average, larger clusters

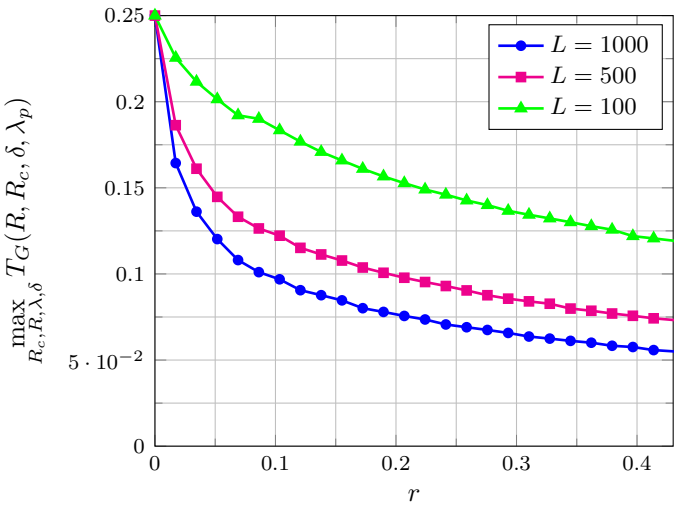

Figure 7.6: Inner bound (7.47) given by $(\Phi_{HC}, \mathcal{F}^*)$ to the Global trade-off region (7.21). T_G is the global fraction requests served through D2D for an average rate constraint $\bar{R} \geq r$. L is the library size. $\lambda_r = 0.003$, $\lambda_u = 4\lambda_r$, $M = 6$, $\gamma = 0.6$, $\alpha = 4$. $\varepsilon = 0.05$.

to find matches, increasing the chance of failed transmissions through path loss. This can be mitigated with a bigger cluster separation $\delta \geq 2R_c$, which reduces the density of clusters, and hence the average level of interference, but may negatively impact the density of served requests, because the clusters cover a smaller fraction of the network. A pair (R_c, δ) should be chosen to balance these effects; although not plotted, simulations show that in this setup, the optimal value was always $\delta = 2R_c$. In addition, by using Lemma 7.2 and the density of the hard core process (7.9), we can find an upper bound for the global metric of this protocol:

$$T_G(\mathcal{F}^*, R, R_c, \delta) \leq (1 - e^{-\lambda\delta^2})\mathbb{P}(\mathcal{M})\frac{R_c^2}{\delta} \leq \frac{1}{4}, \tag{7.48}$$

for any R and hence any average rate constraint. In the last inequality we used that $\delta \geq 2R_c$. Also, when the rate is unconstrained (take $r = 0$ in (7.47)) then by taking $\delta = 2R_c$, a large value of λ (which saturates the cluster density for a given R_c), and a rate $R \to 0$, the bound is achieved, as the plot reflects. If the global metric is interpreted as fraction of users not requiring a download from the BS, the plot shows that even through this simple clustering strategy, a reasonable number of overall demands could be served through D2D.

The global trade-off does not guarantee a certain percentage of service within a cluster, which may be important in some scenarios: there may be many clusters with a small percentage of served requests (small T_L) which results in large overall benefits (large T_G). In Fig. 7.7 we plot the inner bound of the local metric trade-off region (7.22) achievable by $(\Phi_{HC}, \mathcal{F}^*)$,

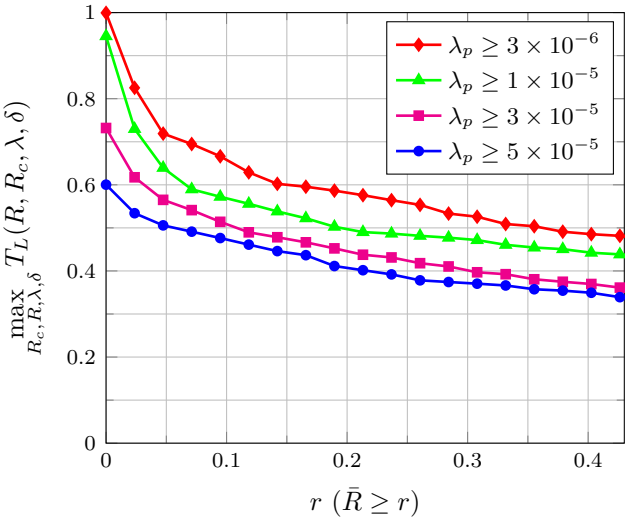

Figure 7.7: Inner bound (7.49) given by $(\Phi_{HC}, \mathcal{F}^*)$ to the local trade-off region (7.22). T_C is the average fraction of served requests per cluster through D2D, for a given average rate $\bar{R}(R_c, R)$ and density of clusters λ_p constraints. $\lambda_r = 0.003 = \lambda_u/4$, $L = 500$, $M = 6$, $\gamma = 0.6$, $\alpha = 4$. $\varepsilon = 0.05$.

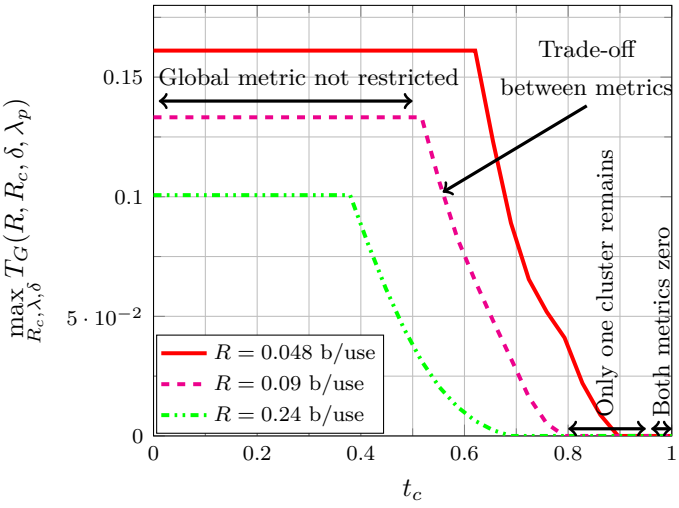

Figure 7.8: Inner bound (7.50) given by $(\Phi_{HC}, \mathcal{F}^*)$ to the Local-Global trade-off region (7.23). T_G is the global fraction of requests served through D2D, while $T_L \geq t_c$ is the fraction of requests served per cluster. R is attempted rate. $\lambda_r = 0.003 = \lambda_u/4$, $L = 500$, $M = 6$, $\gamma = 0.6$, $\alpha = 4$. $\varepsilon = 0.05$.

which maximizes the mean ratio of served requests:

$$\max_{R_c,R,\lambda_p,\delta} T_L(R, R_c, \delta, \lambda_p) \text{ subject to } \bar{R}(R, R_c, \lambda, \delta) \geq r \text{ and } \lambda_p(\delta, \lambda) \geq \lambda_t. \tag{7.49}$$

The fraction of per cluster served requests, which is not bounded by the coverage percentage of cluster process, can be much larger according to setup parameters and the value of the restrictions. As in the global trade-off, we can use Lemma 7.2 to find the value of T_L when the rate is unconstrained ($r = 0$), or as $r \to 0$. We use that:

$$\lambda_p = \frac{(1 - e^{-\lambda\pi\delta^2})}{\pi\delta^2} \leq \frac{1}{4\pi R_c^2},$$

with equality obtained as $\lambda \to \infty$ with $2R_c = \delta$. On the other hand, as $r \to 0$ we have that $T_L(\mathcal{F}, R) \to p_{\mathcal{M}}$. Since $p_{\mathcal{M}}$ increases with R_c and λ_p decreases, the maximization of T_L as $r \to 0$ is achieved by finding the maximum value of $R_c = \delta/2$ which verifies the constraint on λ_p.

Finally, in Fig. 7.8 we plot the local-global metrics trade-off region (7.23) inner bound given by $(\Phi_{HC}, \mathcal{F}^*)$, which is:

$$\max_{R_c,\lambda,\delta} T_G(R, R_c, \delta, \lambda_p) \text{ subject to } T_L(R, R_c, \delta, \lambda_p) \geq t_c. \tag{7.50}$$

We see that if the constraint on local served requests t_c is small enough, then the global metric can be maximized without restriction, and in this regime, the local metric can be set as needed. A smaller local metric implies there is a large number of clusters with a low fraction of served requests, while a larger local metric implies fewer, larger clusters with a higher level of service. If t_c is larger, both constraints cannot be satisfied at the same time, and there is a trade-off between the metrics, which depends on the selected setup. If t_c is even larger, then the maximal density of served requests becomes negligible, which implies that only the cluster at the origin remains. Finally, if the local constraint is too large it cannot be satisfied for any parameters. These observations could be a consequence of the choice of the parent process, and may be different in other cases.

7.5 Summary and discussion

We described a framework which can be used to evaluate the theoretical performance attainable through D2D. In this framework we considered a simple communication strategy, which gives an inner bound to the performance trade-off regions introduced. The analysis shows that, even through our simple communication protocol, a reasonable number of requests could be served through a distributed caching policy without a dedicated caching infrastructure, reducing the downlink traffic on the cellular band and a reduced load on the backhaul.

It is worth mentioning that this framework is quite flexible; other strategies could be considered under this framework, e.g., the behavior of network performance with the parameters

of the caching policy may also be studied, or other caching policies can be considered. A different communication strategy, involving multiple transmitters or other trasmitter-selection algorithms could also be investigated. The interference analysis carried out may be useful in other setups, not limited to D2D.

Chapter 8

Conclusions

The main goal of this work was to study aspects of cooperative communications in the context of large wireless networks, through models in which their principal characteristics could be considered, mainly, a large number of nodes, their unpredictable spatial distribution, and the broadcast nature of the wireless medium. This was motivated by the fact that wireless systems have increased in numbers and in quality requirements, and these increasing demands require new approaches, both structurally and in terms of node interaction and resource optimization. The goal was also to consider models in which rather general conclusions, regarding the network and its performance characteristics, could be drawn. We wanted to obtain, if possible, relatively simple and closed form solutions, meaning by this, expressions which could be evaluated without large-scale Monte Carlo simulations. Stochastic geometry models had shown to be a natural tool for this task which is why we pursued this approach.

We started by considering the full-duplex relay channel under the effect of the interference generated by a uniform distribution of interfering transmitters, modeled by a PPP. We considered this mainly because the relay channel is one of the simplest and perhaps more natural cooperative scenarios. This analysis served as a starting point for studying how traditional coding schemes, which have been developed and studied in the context of classical information theory [4, 23], should be considered in the context of outage-based point process models. The essence of this treatment is given by the fact that the massive interference channel generated in the wireless network is treated as spatially localized communication channels under the presence of correlated noise at the receivers, which is affected by what the transmitters in the network do. The key issues to take into account appeared through this analysis, mainly, the correlation of interference time signals, and its impact on the achievable rates, which are determined by the decoding procedure at the receivers. This implies, for example, that not all DF protocols will have the same outage events and, hence, the same OP, a fact which is many times not precisely considered in the literature. Understanding this correlation and

its effects was one of the main aspects to understand, since under the usual assumption of uncorrelated background noises, or in point-to-point communications in spatial models it is not needed. Another aspect is the correlation among interference powers, which plays an important role through the ubiquitous SINR. Hopefully, this analysis will serve as a starting point for the analysis of more complex network systems and to study the interactions among nodes in large wireless networks, involving different coding and medium-access schemes and network infrastructure.

After considering only one relay channel in the presence of interference, we wanted to study the balance between the benefits of cooperation and the increase interference it generates. We approached this through a simple randomized scheme in which every user can possibly use a relay, which is activated randomly. In an attempt to determine the best performance attainable through this scheme, we obtained the optimal relay activation strategy for this network, which minimized the OP. It was interesting to find that all the relays should be on or off at the same time for the whole network, at least in the regime in which the OP is reasonably small. A justification for this is perhaps that the interference levels grow uniformly in all receivers and hence, at some point attempting to use the relays becomes impractical. More realistic communication schemes should employ the local CSI available at the users to determine whether to use the relay or not. We have performed a simple analysis and drawn some conclusions, about this when the quality of the source-relay channel is used to activate the relay [86, v1], and we have shown here through simulation that this does not seem to bring substantial improvements with respect to the random scheme. A potential improvement could be obtained using more sophisticated cooperative transmission schemes which could take into account the impairments generated by the nearby interferers [76] (which introduce by far the most harmful interference). Basically, this could consist on decoding the messages sent by strong nearby interferers first, then subtracting them from the received signal and finally attempting to decode the desired message. In such situation, besides the intrinsic benefits of cooperation, the smart use of the aggregate interference introduced in part by the cooperating nodes, could ameliorate its harmful effect on the overall network.

In the third phase of this work we attempted to consider cooperation from a different point of view, that is, not only cooperation by joint transmissions, but also through the sharing of content which is locally stored at the users. This setup is interesting because it contained several aspects aside of the problem of successful transmissions. On one hand, the availability, or not, of a requested content at the users was to be considered; otherwise, the success probability of a transmission may be very large, and the probability of a transmission taking place may be very low. On the other hand, to study the success probability of a transmission when the content is locally available, it is necessary to consider which other

transmissions will take place nearby. This requires to define a scheduling policy or medium access strategy in the network. Finally, considering that the transmissions are scheduled in a time-block allows to consider the problem of the time resource distribution between the users, which is generally not considered. In order to analyze a strategy for D2D file exchanges we tried to provide a framework in which a notion of optimality could be defined, that is, to define metrics which had a clear meaning, and under which the analyzed strategy could be compared to a *best* strategy. Finally, we believe that performing these type of analyses, in which several different aspects are considered at the same time may yield interesting general insight about network performance. Otherwise, considering specific problems in great detail may yield important specific results, but at the same time may disregard some key aspects of the problem.

8.1 Future directions

In the context of cooperative transmissions, such as in relaying, in which a coding scheme or protocol is divided into blocks, it would be interesting to consider the effects of the errors between blocks and the correlation between these errors caused by the spatial distribution of the nodes. For example, in the context of full-duplex DF, the relay attempts to decode the transmission from the source in the first block and is active in the second block if it was able to decode in the first one. In a network in which all the users use relays, the interference during the first block would be caused only by the sources, while in the second block it would be caused by the same sources and the relays which managed to decode in the first block. The sources whose relays did not decode in the first block would continue using DT. With the same idea it would be interesting to consider other scenarios such as unicasting and multicasting, in which there are several phases and the users who are able to decode in the first phase then take a cooperative role in the next one; since many interferers will be active in several phases, the correlation between the interferences has to be considered. A first step in this direction was taken in our conference paper [87] in which we analyze the OP of a two-phase unicast strategy introduced in [88].

The analysis of D2D communications introduces several possible ideas to continue, not limited to the D2D scenario exclusively. It would be interesting to study other file caching policies, transmission scheduling strategies and also possibly cooperative transmission schemes such as multicasting. A simpler scheduling strategy, requiring less coordination, would be to consider a fixed number of slots per block and an ALOHA access scheme for the users. Also, since users in a cluster will be close, it would also be interesting to explore the possibility of simultaneous transmissions and interference cancellation [89]. It would also be interesting

to consider a user-centric model, in which a user looks for the video in its vicinity and not inside a cluster. This is because looking in a cluster may imply that a user at the edge is unable to receive a file from a user who is very close but outside of the cluster. The choice of one structure over the other should consider the degree of coordination that may be expected or required in such a network. By this we mean how the interaction between the users and the BS will be: if the BS will pair the users and coordinate the transmissions among the users, or if users will have a higher degree of independence and establish links and transmit on their own accord. This model could also be used to study the benefits that would be obtained by deploying dedicated local file servers to distribute videos in the clusters rather than have the users exchange the videos. In this context, the server may have a wired backhaul or simply overhear transmissions to other users and opportunistically cache files or parts of files, thus facilitating their deployment. On the other hand, in the context of general networks, we believe the study of the time-varying interference presented here deserves further attention, mainly, to study which are the factor which may generate an asynchronism in the transmissions and the interference between different users, and how this interference may impact performance, that is, if it is worth to consider this into the performance of the users or not. For example, in Chapter 7, if it is necessary to consider the time-variations of the interference in the performance analysis, or considering a simpler interference model is enough for an analysis. Another possible direction is studying new point process models which can be used in place of the PPP to model communication scenarios in which the uniformity of nodes is not reasonable. This should be done both in terms of using new processes with, for example, repulsive, attractive or hard-core nature, and in terms of developing and justifying suitable approximations such as the one suggested in [85], that help the derivation of closed form results.

Appendix A

A brief summary on slow-fading non-selective wireless models

Wireless channels are generally modeled as linear time-varying systems, represented by a (lowpass equivalent) time-varying impulse response $c(t, \xi)$, such that, given a (lowpass equivalent) input signal $x(t)$ produces an output $y(t)$:

$$y(t) = \int x(t-\xi)c(t,\xi)d\xi. \tag{A.1}$$

The function $c(t, \xi)$ represents the response at time t of an input impulse ξ time units into the past. Using Fourier transform relations it can be shown that the output $y(t)$ can be equivalently written as [90]:

$$y(t) = \iint x(t-\xi)e^{j2\pi\nu t}C(\xi,\nu)d\nu d\xi, \tag{A.2}$$

where $C(\xi, \nu)$ is the Fourier transform of $c(t, \tau)$ with respect to τ. This implies that the output of a time varying linear system is subject to attenuation and both a scattering in time and frequency. The time scattering property is also present in time invariant systems, while frequency or Doppler spread is caused by the time variations of the channel, implying that more spread is caused by more frequent time changes.

The time-varying impulse response is usually assumed to be a random process in the t variable, whose general characterization is quite involved. Typical assumptions consider the channel to be wide sense stationary in t, in order to characterize the response of the system in terms of its autocorrelation function:

$$R_c(t, \Delta t; \xi_1, \xi_2) = \mathbb{E}\left[c^*(t, \xi_1)c(t + \Delta t, \xi_2)\right] \equiv R_c(\Delta t; \xi_1, \xi_2), \tag{A.3}$$

which is not a function of t according to the stationarity assumption. The hypothesis of *uncorrelated scattering* assumes that different delays provide uncorrelated phase shifts and

attenuations, which implies that:

$$R_c(\Delta t; \xi_1, \xi_2) = R_c(\Delta t, \xi_1)\delta(\xi_2 - \xi_1). \tag{A.4}$$

The Fourier transform with respect to the time delay variable:

$$S(\lambda, \xi) = \int R_c(\Delta t, \xi) e^{-2\pi j \lambda \Delta t} d\Delta t, \tag{A.5}$$

is referred to as the [24] *scattering function* of the channel, which gives a measure of the average power of the channel as a function of the time delay ξ and the Doppler frequency λ. Integrating over each variable separately we can define:

$$S_c(\xi) = \int S(\tau, \lambda) d\lambda \quad \text{Multipath intensity profile} \tag{A.6}$$

$$S_c(\lambda) = \int S(\tau, \lambda) d\tau \quad \text{Doppler power spectrum} \tag{A.7}$$

The interval in which the multipath intensity profile is essentially nonzero, which is known as the multipath spread T_m gives an indication of time it takes for the channel output to be uncorrelated from its input, that is, it is an indication of the average length of the response of the channel. The interval in which the Doppler power spectrum is essentially nonzero, which is known as the Doppler spread of the channel B_d, indicates *how often* the channel changes in time. For this reason, the pair (T_m, B_d) are useful in assessing the behavior of the output of the channel when an input with certain characteristics is present. We can also define the reciprocals and time/frequency counterparts of these quantities: the *channel coherence time* $T_{\text{coh}} = B_d^{-1}$ and the *channel coherence bandwidth*, $B_{\text{coh}} = T_m^{-1}$.

The multipath spread T_m gives an indication of the time it takes for the channel output to be uncorrelated from its input. Given an input with bandwidth W, the condition $W \ll B_{\text{coh}}$ implies that all input frequencies are all equally affected, that is, the *frequency non-selective* or *flat fading*. This implies that almost all multipath components have similar delays, which simplifies (A.1) to a multiplicative channel model (noise not considered):

$$y(t) = \beta(t) e^{j\theta(t)} x(t), \tag{A.8}$$

where $\beta(t)$ is the envelope and $\theta(t)$ represents the phase of the lowpass equivalent channel response. On the other hand, if the duration of the transmitted symbols T_s satisfies $T_s \ll T_{\text{coh}}$, then the channel is said to be *slowly fading* and the multiplicative factor in (A.8) can be considered constant. When these conditions are not met we are in the presence of *frequency selective* or *fast fading* channels (see [24] for details). There are several statistical models for the distribution of the fading coefficients for slow-fading non-selective channels. If the there are a large number of scatterers which contribute to the received signal, the application of

the central limit theorem leads to a model in which the the phase θ is uniform in $[0, 2\pi)$ and independent of the amplitude, which is Rayleigh distributed.

Additionally, there other factors, additional to fading, which can be included in wireless channel models. For example, the previous fading model does not consider the attenuation of signals or *path loss* caused by the spatial separation between the source and the destination of a transmission. Signal propagation is a complex phenomenon to model and finding a general model is quite involved (see [3] for a description of several of these models). A simple model for this attenuation is to assume that a signal $x(t)$ transmitted to a point at distance d is received as:

$$y(t) = \sqrt{K} \left(\frac{d_0}{d} \right)^{\frac{\alpha}{2}} x(t), \tag{A.9}$$

where K is a unitless constant that depends on the antenna characteristics and the average channel attenuation, d_0 is a reference distance and α (generally in the range $(2,\ 6)$) is the path loss exponent. In this work we shall consider that the users' positions do not change during a transmission, that is, the time scale of a transmission is much smaller than the time-scale in which a users may move in space. This implies that path-loss will be constant during a transmission and thus, including it in our slow-fading frequency non-selective model, we have that a transmission of $x(t)$ at a distance d is received as:

$$y(t) = d^{-\frac{\alpha}{2}} \beta e^{j\theta} x(t), \tag{A.10}$$

where we have grouped all the constants into β. A third phenomenon, which we will not consider in our propagation model, is that of *shadowing*, which is related to the random attenuation of signals due to blockage of large objects.

Appendix B

Proofs of Chapter 5

B.1 Rates of decode-and-forward and compress-and-forward

This section provides the basic ideas to evaluate the mutual information appearing in the rates of CF and DF for Gaussian channels. Conditioned on $\tilde{\Phi}$, the signals received at the relay and the destination in each channel use are:

$$Y_r = g_{sr} X_s + \tilde{Z}_r \tag{B.1}$$

$$Y_d = g_{sd} X_s + g_{rd} X_r + \tilde{Z}_d, \tag{B.2}$$

where the channel coefficients g_{sr}, g_{sd}, g_{rd} contain the fading and the path loss. The noises $\tilde{Z}_r$, $\tilde{Z}_d$ are zero-mean CCSG random variables with correlation matrix:

$$Q_N = \begin{bmatrix} I_r & \rho_N \sqrt{I_r I_d} \\ \rho_N^* \sqrt{I_r I_d} & I_d \end{bmatrix},$$

and independent of X_s and X_r, which are also complex Gaussian with correlation matrix:

$$Q = \begin{bmatrix} P_s & \rho \sqrt{P_s P_r} \\ \rho^* \sqrt{P_s P_r} & P_r \end{bmatrix}.$$

In our setup we have $P_s = P_r = 1$.

In the case of CF, in particular we have that $\rho = 0$ and:

$$\hat{Y}_r = Y_r + Z_c, \tag{B.3}$$

where Z_c is zero-mean CCSG noise of variance n_c independent of everything else. The entropy of a CCSG random vector V of covariance matrix $\tilde{Q}$ is [4]:

$$H(V) = \log_2 \det(\pi e \tilde{Q}). \tag{B.4}$$

Using this result, together with the chain rules for mutual information and entropy [4], all the mutual informations can be found. Below we give and example for DF and another for CF.

B.1.1 Calcultion of the rate in $\mathcal{A}_{DF}(R,\rho)$

We write:

$$I(X_s;Y_r|X_r) = H(X_s|X_r) - H(X_s|Y_r,X_r). \tag{B.5}$$

The triple (X_s, X_r, Z_r) are zero-mean CCSG random variables with covariance matrix:

$$\Sigma_1 = \begin{bmatrix} P_s & \rho\sqrt{P_sP_r} & 0 \\ \rho^*\sqrt{P_sP_r} & P_r & 0 \\ 0 & 0 & I_r \end{bmatrix}.$$

We may write:

$$\begin{bmatrix} X_s \\ X_r \\ Y_r \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ g_{sr} & 0 & 1 \end{bmatrix} \begin{bmatrix} X_s \\ X_r \\ \tilde{Z}_r \end{bmatrix} \triangleq A_1 \begin{bmatrix} X_s \\ X_r \\ \tilde{Z}_r \end{bmatrix}.$$

Since A_1 is Hermitian we have that (X_s, X_r, Y_r) are CCSG random variables with covariance matrix Σ_2 such that:

$$\det(\Sigma_2) = \det(A_1)\det(\Sigma_1)\det(A_1^H) = \det(\Sigma_1) = I_r \det(Q).$$

In addition, (Y_r, X_r) are zero-mean CCSG random variables with covariance matrix:

$$\Sigma_3 = \begin{bmatrix} P_r & (g_{sr}\rho\sqrt{P_sP_r})^* \\ g_{sr}\rho\sqrt{P_sP_r} & I_r + |g_{sr}|^2 P_s \end{bmatrix}. \tag{B.6}$$

With this and using (B.4) we have that the first term on the right side of (B.5) is:

$$\begin{aligned} H(X_s|X_r) &= H(X_s,X_r) - H(X_r) \\ &= \log_2\det(\pi e Q) - \log_2(\pi e P_r) \\ &= \log_2(\pi e \det(Q)/P_r) \end{aligned} \tag{B.7}$$

The second term on the right side of (B.5) is:

$$\begin{aligned} H(X_s|Y_r,X_r) &= H(X_s,X_r,Y_r) - H(Y_r,X_r) \\ &= \log_2\det(\pi e\Sigma_2) - \log_2\det(\pi e\Sigma_3) \\ &= \log_2(\pi e I_r \det(Q)\det(\Sigma_3^{-1})) \end{aligned} \tag{B.8}$$

Replacing (B.7) and (B.8) in (B.5) we have:

$$\begin{aligned} I(X_s;Y_r|X_r) &= \log_2(\pi e \det(Q)/P_r) - \log_2(\pi e N_r \det(Q)\det(\Sigma_3^{-1})). \\ &= \log_2\left(1+|g_{sr}|^2\frac{P_s}{I_r}\left(1-|\rho|^2\right)\right). \end{aligned} \tag{B.9}$$

B.1.2 Evaluation of $I(X_s;\hat{Y}_r,Y_d|X_r)$ for compress-and-forward

We have that:

$$I(X_s;\hat{Y}_r,Y_d|X_r) = H(X_s|X_r) - H(X_s|\hat{Y}_r,Y_d,X_r) \tag{B.10}$$

$$= H(X_s) + H(\hat{Y}_r,Y_d,X_r) - H(X_s,\hat{Y}_r,Y_d,X_r). \tag{B.11}$$

For the second term we have:

$$\begin{bmatrix} X_r \\ \hat{Y}_r \\ Y_d \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ g_{sr} & 0 & 1 & 0 & 1 \\ g_{sd} & g_{rd} & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_s \\ X_r \\ \tilde{Z}_r \\ \tilde{Z}_d \\ Z_c \end{bmatrix} \tag{B.12}$$

so the correlation matrix of the zero-mean CCSG random variables $(X_r,\hat{Y}_r,Y_d)$ is:

$$\Sigma_4 = \begin{bmatrix} P_r & 0 & P_r g_{rd}^* \\ 0 & I_r + n_c + P_s|g_{sr}|^2 & \rho_N\sqrt{I_r I_d} + P_s g_{sr} g_{sd}^* \\ P_r g_{rd} & \rho_N^*\sqrt{I_r I_d} + P_s g_{sd} g_{sr}^* & I_d + P_s|g_{sd}|^2 + P_r|g_{rd}|^2 \end{bmatrix}, \tag{B.13}$$

from where we get:

$$H(\hat{Y}_r,Y_d,X_r) = \log_2(\det(\Sigma_4)). \tag{B.14}$$

Para el tercer término el cambio es:

$$\begin{bmatrix} X_s \\ X_r \\ \hat{Y}_r \\ Y_d \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ g_{sr} & 0 & 1 & 0 & 1 \\ g_{sd} & g_{rd} & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_s \\ X_r \\ Z_r \\ Z_d \\ Z_c \end{bmatrix}, \tag{B.15}$$

and the resulting correlation matrix is:

$$\Sigma_5 = \begin{bmatrix} P_s & 0 & P_s g_{sr}^* & P_s g_{sd}^* \\ 0 & P_r & 0 & P_r g_{rd}^* \\ P_s g_{sr} & 0 & n_c + I_r + P_s|g_{sr}|^2 & P_s g_{sr} g_{sd}^* \\ P_s g_{sd} & P_s h_{rd} & P_s g_{sd} g_{sr}^* & I_d + P_r|g_{rd}|^2 + P_s|g_{sd}|^2 \end{bmatrix} \tag{B.16}$$

We therefore have:

$$I(X_s;\hat{Y}_r,Y_d|X_r)=\log_2\left(\frac{P_s\det(\Sigma_4)}{\det(\Sigma_5)}\right). \tag{B.17}$$

$$=\log_2\left(1+P_s\frac{|g_{sd}|^2(n_c+I_r)+I_d|g_{sr}|^2-2\sqrt{I_rI_d}\Re\{\rho_N g_{sd}g_{sr}^*\}}{(n_c+I_r)I_d-|\rho_N|^2I_rI_d}\right). \tag{B.18}$$

B.2 Proof of Lemma 5.1

Starting from (5.8) we factorize the integrand as:

$$1-\frac{1}{(\omega_1 l(x,z_1)+1)(\omega_2 l(x,z_2)+1)}=\frac{1}{1+\frac{1}{\omega_1 l(x,z_1)}}+\frac{1}{1+\frac{1}{\omega_2 l(x,z_2)}} - \frac{1}{(1+\frac{1}{\omega_1 l(x,z_1)})(1+\frac{1}{\omega_2 l(x,z_2)})}. \tag{B.19}$$

If we now integrate this sum, the first two terms can be found in closed form:

$$\int_{\mathbb{R}^2}\frac{1}{1+(\omega_1||x-r||^{-\alpha})^{-1}}dx=\tilde{C}\omega_1^{2/\alpha}, \tag{B.20}$$

an integral which also appears in the case of a DT [10]. The third term is (5.11) after replacing the path loss expression with its expression $l(x,d)=||x-d||^{-\alpha}$, $l(x,r)=||x-r||^{-\alpha}$.

B.3 Proof of Theorem 5.2

In what follows we omit the dependence of the outage event on (R,ρ) and write T instead of $T(R)$. We start by (5.49), for which we have:

$$\mathbb{P}\left(\mathcal{A}_{DF}\cap\mathcal{A}_{DT}^c\right)=\mathbb{E}_{\tilde{\Phi}}\left[\mathbb{P}\left(|h_{sr}|^2<\frac{TI_r}{\mu_3},|h_{sd}|^2\geq\frac{TI_d}{l_{sd}}\middle|\tilde{\Phi}\right)\right],$$

$$=\mathbb{E}_{\tilde{\Phi}}\left[\left(1-e^{-\frac{TI_r}{\mu_3}}\right)e^{-\frac{TI_d}{l_{sd}}}\right], \tag{B.21}$$

$$=\mathbb{E}_{\tilde{\Phi}}\left[e^{-\frac{TI_d}{l_{sd}}}-e^{-\left(\frac{TI_d}{l_{sd}}+\frac{TI_r}{\mu_3}\right)}\right] \tag{B.22}$$

where we considered that the power fading coefficients are independent exponential unit-mean random variables. Applying the definition of the Laplace transform (5.5) we have (5.49). To find (5.44) we define:

$$Z\triangleq|h_{sd}|^2l_{sd}+|h_{rd}|^2l_{rd}+2\sqrt{l_{sd}l_{rd}}\Re(\rho h_{sd}h_{rd}^*), \tag{B.23}$$

so:

$$\mathbb{P}\left(\mathcal{A}_{DF}^c\cap\mathcal{B}_{DF}^c\right)=\mathbb{E}_{\tilde{\Phi}}\left[\mathbb{P}\left(|h_{sr}|^2\geq\frac{TI_r}{\mu_3},Z\geq TI_d\middle|\tilde{\Phi}\right)\right],$$

$$=\mathbb{E}_{\tilde{\Phi}}\left[e^{-\frac{TI_r}{\mu_3}}\bar{F}_Z(TI_d)\right], \tag{B.24}$$

where $\bar{F}_Z(\cdot)$ is the complementary cumulative distribution function of Z and μ_3 is given by (5.47). The complementary cumulative distribution function Z is:

$$\bar{F}_Z(u) = \begin{cases} \frac{\mu_2 e^{-u/\mu_2} - \mu_1 e^{-u/\mu_1}}{\mu_2 - \mu_1} & \mu_1 \neq \mu_2 \\ (1 + u/\mu_1) e^{-u/\mu_1} & \mu_1 = \mu_2 \end{cases} \tag{B.25}$$

To see this, notice that Z can be written as:

$$Z = q^H \begin{bmatrix} l_{sd} & \sqrt{l_{sd} l_{rd}}\rho \\ \sqrt{l_{sd} l_{rd}}\rho^* & l_{rd} \end{bmatrix} q \triangleq q^H Q q, \tag{B.26}$$

where $q = [h_{sd},\ h_{rd}]^T$, is a zero-mean complex circularly symmetric Gaussian vector with identity covariance matrix and Q is positive definite. Diagonalizing Q we can write $Z = \mu_1 |w_1|^2 + \mu_2 |w_2|^2$, where μ_1 and μ_2 are the eigenvalues of Q, given by (5.45) and (5.46), and $|w_1|^2$ and $|w_2|^2$ are unit mean exponential variables, the same as $|h_{sd}|^2$ and $|h_{rd}|^2$. Applying a straightforward change of variables from $\{|w_1|^2, |w_2|^2\}$ to Z we obtain (B.25). Notice that when $\rho = 0$ and $||d - r|| = D$, μ_1 and μ_2 coincide so Z follows a Gamma distribution with 2 degrees of freedom, which accounts for the case $\mu_1 = \mu_2$. Now, replacing (B.25) in (B.24) and using the definition of the Laplace transform (5.5) we obtain (5.44). To obtain (5.48) we replace (B.25) in (B.24) and use the fact that:

$$\frac{d\mathbb{E}_{\tilde{\Phi}}\left[e^{-(\omega_1 I_d + \omega_2 I_r)}\right]}{d\omega_1} = \mathbb{E}_{\tilde{\Phi}}\left[-I_d e^{-(\omega_1 I_d + \omega_2 I_r)}\right]. \tag{B.27}$$

B.4 Proof of Lemma 5.3

For this proof we write T instead of $T(R)$, for short. We start from the expression of the Laplace transform (5.8) which we repeat here for simplicity:

$$\mathcal{L}_{I_d, I_r}(\omega_1, \omega_2) = \exp\left\{-\lambda \int_{\mathbb{R}^2} \left[1 - \frac{1}{(1 + \omega_1 l(x,d))(1 + \omega_2 l(x,r))}\right] dx\right\}. \tag{B.28}$$

Using the big-O notation, this Laplace transform can be written as:

$$\mathcal{L}_{I_d, I_r}(\omega_1, \omega_2) = 1 - \lambda \int_{\mathbb{R}^2} \left[1 - \frac{1}{(1 + \omega_1 l(x,d))(1 + \omega_2 l(x,r))}\right] dx + O(\lambda^2), \tag{B.29}$$

and hence, using this equation and Theorem 5.2 we can write:

$$\mathbb{P}\left(\mathcal{A}_{DF}(R,\rho) \cap \mathcal{A}_{DT}^c\right) = \mathcal{L}_{I_d}\left(\frac{T}{l_{sd}}\right) - \mathcal{L}_{I_d, I_r}\left(\frac{T}{l_{sd}}, \frac{T}{\mu_3}\right) \tag{B.30}$$

$$= \lambda \int_{\mathbb{R}^2} \frac{l_{sd}}{l_{sd} + T l(x,d)} \left(1 - \frac{\mu_3}{\mu_3 + T l(x,r)}\right) dx + O(\lambda^2), \tag{B.31}$$

and:

$$\mathbb{P}\left(\mathcal{A}_{DF}^{c}(R,\rho)\cap\mathcal{B}_{DF}^{c}(R,\rho)\right)= 1-\lambda\int_{\mathbb{R}^2}\left\{1-\frac{\mu_3(\mu_2-\mu_1)^{-1}}{\mu_3+Tl(x,r)}\left[\frac{\mu_2^2}{\mu_2+Tl(x,d)}-\frac{\mu_1}{\mu_1+Tl(x,d)}\right]\right\}dx+O(\lambda^2), \quad \text{(B.32)}$$

where we write T instead of $T(R)$. Using (B.31) and (B.32) in (5.43) the OP can be written as:

$$P_{\text{out,DF}}=\lambda\int_{\mathbb{R}^2}\left(1-\frac{l_{sd}}{l_{sd}+Tl(x,d)}\right)dx-\lambda\int_{\mathbb{R}^2}\left\{\frac{\mu_3}{\mu_3+Tl(x,r)} \times\left[\frac{\mu_2^2(\mu_2-\mu_1)^{-1}}{\mu_2+Tl(x,d)}-\frac{\mu_1^2(\mu_2-\mu_1)^{-1}}{\mu_1+Tl(x,d)}-\frac{l_{sd}}{l_{sd}+Tl(x,d)}\right]\right\}dx+O(\lambda^2). \quad \text{(B.33)}$$

The first term in (B.33) does not depend on $|\rho|$ so we can ignore it. To conclude the proof, we check that the integrand in the second integral is decreasing in $|\rho|$, which implies that the OP is increasing. To do this, we write μ_1 and μ_2 as:

$$\mu_1=a-b(\rho), \quad \text{(B.34)}$$

$$\mu_2=a+b(\rho). \quad \text{(B.35)}$$

where $a=\frac{1}{2}(l_{sd}+l_{rd})$ does not depend on ρ and $b(\rho)=\frac{1}{2}\left[(l_{sd}-l_{rd})^2-4l_{sd}l_{rd}|\rho|^2\right]^{\frac{1}{2}}$ does. We also define, for $u\geq 0$, the following function:

$$\varphi(u)\triangleq\frac{(a+u)^2}{u(a+u+Tl(x,d))}. \quad \text{(B.36)}$$

Using a, b and φ we can write the second integrand in (B.33) as:

$$\frac{\mu_3}{\mu_3+Tl(x,r)}\left[\frac{\mu_2^2(\mu_2-\mu_1)^{-1}}{\mu_2+Tl(x,d)}-\frac{\mu_1^2(\mu_2-\mu_1)^{-1}}{\mu_1+Tl(x,d)}-\frac{l_{sd}}{l_{sd}+Tl(x,d)}\right]= \frac{\mu_3(|\rho|)}{\mu_3(|\rho|)+Tl(x,r)}\left[\frac{\varphi(b(|\rho|))+\varphi(-b(|\rho|))}{2}-\frac{l_{sd}}{l_{sd}+Tl(x,d)}\right], \quad \text{(B.37)}$$

where we have made explicit all the elements which depend on $|\rho|$. We conclude the proof by showing that the right side of (B.37) is the product of two positive functions which are decreasing in $|\rho|$. Since $\mu_3=l_{sr}\left(1-|\rho|^2\right)$ is decreasing in $|\rho|$ we see that:

$$\frac{\mu_3(|\rho|)}{\mu_3(|\rho|)+Tl(x,r)} \quad \text{(B.38)}$$

is decreasing in $|\rho|$ (and it is also positive). To show that the rest is decreasing, we check that

$$\frac{d(\varphi(b)+\varphi(-b))}{db}=\frac{-4T^2l(x,d)^2b}{(T^2l(x,d)^2+2Tl(x,d)a+a^2-b^2)^2}<0,$$

and since b is increasing in $|\rho|$, then $\varphi(b(|\rho|)) + \varphi(-b(|\rho|))$ is decreasing in $|\rho|$. We conclude taking $|\rho| = 1$ to check that:

$$\frac{\varphi(b(1)) + \varphi(-b(1))}{2} - \frac{l_{sd}}{l_{sd} + Tl(x,d)} > 0,$$

which implies that the function in brackets on the right side of (B.37) is positive.

B.5 Proof of Theorem 5.4

Starting from (5.51), we find the probability of the events $\mathcal{A}_{DF}$ and $\mathcal{B}_{DF}$, by following the steps used to derive (B.22):

$$\mathbb{P}\left(\mathcal{A}_{DF}(R,\rho)\right) = 1 - \mathcal{L}_{I_r}\left(T/(1-|\rho|^2)\right), \tag{B.39}$$

$$\mathbb{P}\left(\mathcal{B}_{DF}(R,\rho)\right) = 1 - \frac{\mu_2 \mathcal{L}_{I_d}\left(T/\mu_2\right) - \mu_1 \mathcal{L}_{I_d}\left(T/\mu_1\right)}{\mu_2 - \mu_1}. \tag{B.40}$$

The Laplace transform can be found in Lemma 5.1 and setting ω_1 or ω_2 to zero as needed [10] $\mathcal{L}_{I_r}(\omega) = \mathcal{L}_{I_d}(\omega) = \exp\{-\lambda \tilde{C} \omega^{2/\alpha}\}$, where $\tilde{C}$ is given by (5.10). Replacing (B.39) and (B.40) in (5.51) and using (5.41) we finish the proof. The proof of the second part follows directly from the fact that $e^{-u} = 1 - u + O(u^2)$ and the definition of $\kappa(R)$ (5.17).

B.6 Proof of Theorem 5.5

In what follows we omit the dependence of the outage event on (R, ϵ). Using the concavity of the logarithm we can bound:

$$R_{SDF} \leq \tilde{R}_{SDF} = \mathsf{C}\left(\frac{|h_{sd}|^2 l_{sd} + (1-\varepsilon)|h_{rd}|^2 l_{rd}}{I_d}\right) \tag{B.41}$$

With this we have $\mathcal{B}_{SDF} \supset \tilde{\mathcal{B}}_{SDF} = \{\tilde{R}_{SDF} < R\}$ and:

$$\mathcal{O}_{SDF} \supset [\mathcal{A}^c_{SDF} \cap \tilde{\mathcal{B}}_{SDF}] \cup [\mathcal{A}_{SDF} \cap \mathcal{A}_{DT}], \tag{B.42}$$

$$= [(\mathcal{A}^c_{SDF} \cap \tilde{\mathcal{B}}^c_{SDF}) \cup (\mathcal{A}_{SDF} \cap \mathcal{A}^c_{DT})]^c \tag{B.43}$$

In the second step we used that $\tilde{\mathcal{B}}_{SDF} \subset \mathcal{A}_{DT}$ and the union is disjoint. The rest of the proof follows along the lines of the proof in Appendix B.3: the first event in (B.43) is the same as that of $\mathcal{B}_{DF}$ by taking $\rho = 0$ and a path loss $(1-\varepsilon)l_{rd}$. The second term is the same as (B.22) with $\rho = 0$ and $T = 2^{R/\epsilon} - 1$.

B.7 Proof of Lemma 5.6

We use the CF rate given by (5.31), and write $R_1 \equiv R_1(0)$ and $R_2 \equiv R_2(0)$:

$$\begin{aligned} R_{CF}(0) - R_{CF}(\rho_N) &= \min\{R_1, R_2\} - \min\{R_1(\rho_N), R_2(\rho_N)\} \\ &\leq \max\{R_1 - R_1(\rho_N), R_2 - R_2(\rho_N)\}. \end{aligned}$$

Now we use that:

$$R_1 - R_1(\rho_N) = \mathsf{C}\left(\frac{I_r}{n_c}(1-|\rho_N|^2)\right) - \mathsf{C}\left(\frac{I_r}{n_c}\right) \tag{B.44}$$

$$= \mathsf{C}\left(-\frac{|\rho_N|^2}{1+\frac{n_c}{I_r}}\right) \leq 0. \tag{B.45}$$

On the other hand, defining:

$$u = \frac{|h_{sd}|\sqrt{l_{sd}}}{\sqrt{I_d}} \qquad v = \frac{|h_{sr}|\sqrt{l_{sr}}}{\sqrt{I_r}} \tag{B.46}$$

we can rewrite and bound the rate $R_2(\rho_N)$ as:

$$\begin{aligned} R_2(\rho_N) &= \mathsf{C}\left(|u|^2 + \frac{|\rho_N|^2|u|^2 + |v|^2 - 2\Re\{\rho_N u v^*\}}{1+\frac{n_c}{I_r} - |\rho_N|^2}\right), \\ &\geq \mathsf{C}\left(|u|^2 + \frac{|\rho_N|^2|u|^2 + |v|^2 - 2|\rho_N||u||v|}{1+\frac{n_c}{I_r} - |\rho_N|^2}\right). \end{aligned} \tag{B.47}$$

Now we are interested in finding a lower bound of (B.47). Using standard analysis techniques it is straightforward to check that in the range $0 \leq \rho_N \leq 1$ this function has a single minimum. We have to consider three regimes:

- $|v| < |u|$: there is a minimum at $|\rho_N| = \frac{|v|}{|u|} < 1$.
- $|u| \leq |v| \leq (1+\frac{n_c}{I_r})|u|$: the function is decreasing ρ_N so there $|\rho| = 1$ gives the smallest rate.
- $(1+\frac{n_c}{I_r})|u| < |v|$: in this case there is a minimum at:

$$|\rho_N| = \frac{(1+\frac{n_c}{I_r})|u|}{|v|} < 1. \tag{B.48}$$

With this analysis, we can show that:

$$R_2(\rho_N) \geq \begin{cases} \mathsf{C}(|u|^2) & \text{if } \frac{|v|}{|u|} < 1, \\ \mathsf{C}\left(|u|^2 + \frac{(|u|-|v|)^2 I_r}{n_c}\right) & \text{if } 1 \leq \frac{|v|}{|u|} \leq (1+\frac{n_c}{I_r}), \\ \mathsf{C}\left(\frac{|v|^2}{1+\frac{n_c}{I_r}}\right) & \text{if } (1+\frac{n_c}{I_r}) < \frac{|v|}{|u|}. \end{cases}$$

Using this fact we conclude by noting that:

$$R_2(0) - R_2(\rho_N) \leq \begin{cases} \mathsf{C}\left(\frac{|v|^2}{|u|^2\left(1+\frac{n_c}{I_r}\right)}\right) \leq 1 & \text{if } \frac{|v|}{|u|(1+\frac{n_c}{I_r})} \leq 1 \\ \mathsf{C}\left(\frac{|u|^2}{|v|^2\left(1+\frac{n_c}{I_r}\right)}\right) \leq 1 & \text{if } \frac{|v|}{|u|(1+\frac{n_c}{I_r})} > 1, \end{cases}$$

B.8 Proof of Theorem 5.7

Once more we write $T \equiv T(R)$. Since, $R_{CF}(\rho_N, n_c) \geq R_{CF}(0, n_c) - 1$, for an attempted rate R, we have:

$$\begin{aligned} P_{\text{out,CF}}(R, n_c) &= \mathbb{P}(R_{CF}(\rho_N, n_c) < R) \\ &\leq \mathbb{P}(R_{CF}(0, n_c) - 1 < R). \end{aligned}$$

For the bound on the probability of $\mathcal{A}_{CF}$ we write:

$$\left\{\frac{|h_{sr}|^2 l_{sr}}{I_r + n_c} < T\right\} = \mathcal{A}_{CF} \cup \left(\mathcal{A}_{CF}^c \cap \left\{\frac{|h_{sr}|^2 l_{sr}}{I_r + n_c} < T\right\}\right).$$

Then we take any $\tilde{N} \in \mathbb{N}$ and use the following inclusion:

$$\mathcal{A}_{CF}^c \cap \left\{\frac{|h_{sr}|^2 l_{sr}}{I_r + n_c} < T\right\} \subseteq \bigcup_{\tilde{n}=0}^{\tilde{N}-1} \left\{\frac{\tilde{n}}{\tilde{N}} T \leq \frac{|h_{sr}|^2 l_{sr}}{I_r + n_c} < \frac{\tilde{n}+1}{\tilde{N}} T \,,\, \frac{|h_{sd}|^2 l_{sd}}{I_d} \geq \frac{\tilde{N} - \tilde{n}}{\tilde{N}} T\right\}.$$

The union in the previous equation is a disjoint coverage of the event on the left side, so we have:

$$\mathbb{P}(\mathcal{A}_{CF}) \leq \mathbb{P}\left(\frac{|h_{sr}|^2 l_{sr}}{I_r + n_c} < T\right) - \sum_{\tilde{n}=0}^{\tilde{N}-1} \mathbb{P}\left(\frac{\tilde{n}}{\tilde{N}} T \leq \frac{|h_{sr}|^2 l_{sr}}{I_r + n_c} < \frac{\tilde{n}+1}{\tilde{N}} T, \frac{|h_{sd}|^2 l_{sd}}{I_d} \geq \frac{\tilde{N} - \tilde{n}}{\tilde{N}} T\right).$$

Now we can condition on the point process and using that the fading coefficients are independent we can write the probabilities in terms of the Laplace transform of the interferences. For the other event, since $n_c > 0$ we have:

$$\begin{aligned} \bar{\mathcal{A}}_{CF}(R+1, n_c, 0) &\subseteq \bar{\mathcal{A}}_{CF}(R+1, 0, 0) \\ &= \left\{\frac{1}{T}\left(\frac{|h_{sr}|^2 l_{sr}}{I_r} + \frac{|h_{sd}|^2 l_{sd}}{I_d}\right) \geq 1\right\}. \end{aligned} \tag{B.49}$$

Noticing that:

$$\mathcal{B}_{CF}(n_c, 0) = \left\{n_c < \frac{I_r I_d}{|h_{rd}|^2}\left(\frac{|h_{sd}|^2}{I_d} + \frac{|h_{sr}|^2}{I_r} + 1\right)\right\}, \tag{B.50}$$

we see that we can use (B.50) with (B.49) to bound:

$$\bar{\mathcal{A}}_{CF}(R+1, n_c, 0) \cap \mathcal{B}_{CF}(n_c, 0) \subseteq \bar{\mathcal{A}}_{CF}(R+1, 0, 0) \cap \mathcal{B}_{CF}(n_c, 0) \tag{B.51}$$

$$\subseteq \left\{\frac{|h_{sr}|^2 l_{sr} I_d + |h_{sd}|^2 l_{sd} I_r}{n_c l_{rd} |\hat{h}_{rd}|^2} > \frac{T}{1+T}\right\}. \tag{B.52}$$

We therefore have:

$$\mathbb{P}\left(\bar{\mathcal{A}}_{CF}(R+1,n_c,0)\cap\mathcal{B}_{CF}(n_c,0)\right)\leq\mathbb{P}\left(\frac{|h_{sr}|^2 l_{sr}I_d+|h_{sd}|^2 l_{sd}I_r}{n_c l_{rd}|h_{rd}|^2}>\frac{T}{1+T}\right)$$
$$=1-\mathbb{E}\left[\mathcal{L}_{I_d,I_r}\left(\frac{(1+T)l_{sr}|h_{sr}|^2}{Tn_c l_{rd}},\frac{(1+T)l_{sd}|h_{sd}|^2}{Tn_c l_{rd}}\right)\right],$$

where the expectation is over $|h_{sr}|^2$ and $|h_{sd}|^2$. To obtain this last expression we first condition on h_{sr}, h_{sd} and the point process, and evaluate the probability with respect to h_{rd}. We then take the expectation with respect to the point process to obtain the joint Laplace transform and finally the expectation with respect to the fading coefficients. All these random elements are independent. By using this bound we avoid working with the product of the interference at the relay and the destination, which complicates the evaluation of $\mathbb{P}(\mathcal{B}_{CF})$ significantly. In addition, when all the distances remain fixed and for sufficiently small λ, the product term will be small and its contribution will not be significant in comparison with the other terms. Now observing the expression of the Laplace transform given in Lemma 5.1 we see that we can lower bound the joint Laplace Transform by removing the function (5.11), which is equivalent to assuming that the interferences are independent and splits the joint Laplace transform into the product of the transforms of the separate interferences.

Appendix C

Proofs of Chapter 6

C.1 Proof of Theorem 6.2

To evaluate the bound (6.19) in terms of the Laplace transforms of the interferences we can follow the steps of Theorem 5.2, to find that:

$$\mathbb{P}\left(\mathcal{A}_{DT}(R)|\varepsilon_0 = 0\right) = 1 - \mathcal{L}_{I_d}\left(T/l_{sd}\right), \tag{C.1}$$

$$\mathbb{P}\left(\mathcal{A}_{DF}(R)|r, \varepsilon_0 = 1\right) = 1 - \mathcal{L}_{I_r}\left(T/l_{sr}\right), \tag{C.2}$$

$$\mathbb{P}\left(\mathcal{B}_{DF}(R)|r, \varepsilon_0 = 1\right) = 1 - \frac{D^\alpha \mathcal{L}_{I_d}\left(T/l_{rd}\right) - ||r-d||^\alpha \mathcal{L}_{I_d}\left(T/l_{sd}\right)}{D^\alpha - ||r-d||^\alpha}. \tag{C.3}$$

Therefore (6.19) becomes:

$$\begin{aligned} P_{\text{out,mix}}(R) \leq\ & (1-p_r)\left(1 - \mathcal{L}_{I_d}\left(T/l_{sd}\right)\right) \\ & + p_r \mathbb{E}_r\left[2 - \mathcal{L}_{I_r}\left(T/l_{sr}\right) + \frac{D^\alpha \mathcal{L}_{I_d}\left(T/l_{rd}\right) - ||r-d||^\alpha \mathcal{L}_{I_d}\left(T/l_{sd}\right)}{D^\alpha - ||r-d||^\alpha}\right]. \end{aligned} \tag{C.4}$$

Using (C.24) from Lemma C.1 (below) we can evaluate all the Laplace transforms to exponential functions. To derive the representation using the big O notation, we define $u = ||r-d||/D$ and write:

$$\frac{D^\alpha \mathcal{L}_{I_d}\left(T/l_{rd}\right)}{D^\alpha - \|r-d\|^\alpha} - \frac{||r-d||^\alpha \mathcal{L}_{I_d}\left(T/l_{sd}\right)}{D^\alpha - \|r-d\|^\alpha} = e^{-\lambda_s \Delta(p_r) D^2} \left(\frac{e^{\lambda_s \Delta(p_r) D^2 (1-u^2)} - u^\alpha}{1-u^\alpha}\right). \tag{C.5}$$

Using that as $\lambda_s \Delta D^2(1-u^2) \to 0$ we have:

$$e^{\lambda_s \Delta D^2 (1-u^2)} = 1 + \lambda_s \Delta D^2 (1-u^2) + O((\lambda_s \Delta D^2 (1-u^2))^2), \tag{C.6}$$

that for $u > 0$:

$$\frac{1-u^2}{1-u^\alpha} \geq 1 + \left(\frac{2}{\alpha} - 1\right) u, \tag{C.7}$$

we find that:

$$\left(\frac{e^{\lambda_s \Delta D^2(1-u^2)}-u^\alpha}{1-u^\alpha}\right) \geq 1+\lambda_s \Delta D^2\left[1+\left(\frac{2}{\alpha}-1\right)u\right]+O\left((\lambda_s \Delta D^2)^2 \frac{\left(1-u^2\right)^2}{|1-u^\alpha|}\right), \quad \text{(C.8)}$$

and taking expectation with respect to u we conclude that:

$$\mathbb{E}_r\left[\frac{e^{\lambda_s \Delta(p_r) D^2(1-u^2)}-u^\alpha}{1-u^\alpha}\right] \geq 1+\lambda_s \Delta(p_r) D^2\left[1+\frac{2-\alpha}{\alpha}\mathbb{E}_r\left[u\right]\right]+O\left((\lambda_s \Delta(p_r) D^2)^2\right). \quad \text{(C.9)}$$

With this, we have found (6.20).

Equation (6.21) can be found by simply evaluating the expectation. Finally, to find (6.22) start by writing:

$$\mathbb{E}_r\left[||r-d||\right]=\int_{\mathbb{R}^2}\frac{1}{2\pi\sigma_{in}^2}||r-d||e^{-\frac{||r||^2}{2\sigma_{in}^2}}dr. \quad \text{(C.10)}$$

Now take $x=r-d$, change to polar coordinates and take $t=u/\sigma$ to obtain:

$$\begin{aligned}
\mathbb{E}\left[||d-r||\right] &= \int_0^\infty \int_0^{2\pi} \frac{1}{2\pi\sigma^2} u^2 e^{-\frac{1}{2\sigma^2}(u^2+2Du\cos(\theta)+D^2)} d\theta du & \text{(C.11)}\\
&= \int_0^\infty \frac{1}{2\pi\sigma^2} u^2 e^{-\frac{u^2+D^2}{2\sigma^2}} \left(\int_0^{2\pi} e^{-\frac{Du\cos(\theta)}{\sigma^2}} d\theta\right) du & \text{(C.12)}\\
&= \int_0^\infty \left(\frac{u}{\sigma}\right)^2 e^{-\frac{u^2+D^2}{2\sigma^2}} I_0\left(\frac{Du}{\sigma^2}\right) du & \text{(C.13)}\\
&= \sigma Q_{2,0}\left(\frac{D}{\sigma},0\right). & \text{(C.14)}
\end{aligned}$$

In the second step we used the definition of the modified Bessel function. For the actual value of $Q_{2,0}(u,0)$ use (91) from [74] which states that:

$$Q_{2,0}(b,ac)=a\left[\int_c^\infty Q_1(b,ax)dx+cQ_1(b,ac)\right], \quad \text{(C.15)}$$

where $Q_1(\cdot,\cdot)$ is the Marcum Q function [91], defined as:

$$Q(a,b)=\int_b^\infty te^{-\frac{t^2+a^2}{2}}I_0(at)dt. \quad \text{(C.16)}$$

Taking $c=0$ in (C.15) we get:

$$Q_{2,0}(b,0)=a\int_0^\infty Q_1(b,ax)dx, \quad \text{(C.17)}$$

and using (60) of [74] which states that:

$$\int_0^\infty Q_1(b,ax)dx=\frac{\sqrt{2\pi}}{4a}e^{-\frac{b^2}{4}}\left[(b^2+2)I_0\left(\frac{b^2}{4}\right)+b^2 I_1\left(\frac{b^2}{4}\right)\right] \quad \text{(C.18)}$$

we obtain the desired result.

To finish we need to prove the bound (6.23), for the case in which $\phi_0 \neq 2\pi$. We first prove that:

$$||r-d|| \leq |||d|| - ||r||| + 2\min\left(||d||, ||r||\right)|\sin\left(\theta/2\right)|, \tag{C.19}$$

where θ is the angle between r and d. We decompose $r-d$ as $r-d = u_1+v_1 = u_2+v_2$ with:

$$u_1 = r - \frac{||r||}{||d||}d \qquad u_2 = \frac{||d||}{||r||}r - d. \tag{C.20}$$

Then we use that $||u_1|| = 2||r||\sin(\theta/2)$, $||u_2|| = 2||d||\sin(\theta/2)$ and $||v_1|| = ||v_2|| = |||r|| - ||d|||$ and the triangle inequality on both decompositions. By taking the expectation on both sides of (C.19) and use the following:

$$\int_0^{2\pi}\left(2\sin\left(\frac{\theta}{2}\right)-1\right)d\theta = 2(4-\pi), \tag{C.21}$$

$$\int uf(u)du = -\frac{1}{2\pi}e^{-\frac{u^2}{2\sigma^2}}, \tag{C.22}$$

$$\int u^2 f(u)du = -\frac{u}{2\pi}e^{-\frac{u^2}{2\sigma^2}} + \frac{\sigma}{\sqrt{8\pi}}\mathrm{erf}\left(\frac{u}{\sqrt{2}\sigma}\right), \tag{C.23}$$

we conclude the proof. It is noted that (C.23) is obtained integrating by parts.

Lemma C.1. *Suppose the marks of an independently marked homogeneous PPP are* $m = (|h_1|^2, |h_2|^2, \varepsilon, k)$*, with* $|h_1|^2$ *and* $|h_2|^2$ *unit mean independent exponential random variables,* ε *a Bernoulli random variable with success probability* p_r*, and* k *a random variable on* $\mathbb{R}^2$*. Let* $f_1(d,x,m) = \left(|h_1|^2 + \varepsilon|h_2|^2\right) l(x+\tau k, d)$ *with the path loss function* $l(x,y) = ||x-y||^{-\alpha}$ *and* $\tau \in [0,1]$*. Then the Laplace transform is:*

$$\mathcal{L}_{I_d}(\omega_1) = \exp\left\{-\lambda_s \tilde{C}\omega_1^{2/\alpha}\left(1+\frac{2p_r}{\alpha}\right)\right\}. \tag{C.24}$$

Proof. Taking $\omega_2 = 0$ in (5.8), writing the expectation with respect to the marks and interchanging the integration order we find that:

$$\mathcal{L}_{I_d}(\omega_1) = \exp\left\{-\lambda_s p_r \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} 1 - \frac{1}{\left[1+\omega_1 l(x+\tau k, d)\right]^2}dx dF_k \right.$$
$$\left. -\lambda_s(1-p_r)\int_{\mathbb{R}^2}\int_{\mathbb{R}^2}\frac{1}{1+(\omega_1 l(x+\tau k,d))^{-1}}dx dF_k\right\}. \tag{C.25}$$

When the integrals with respect to x are computed the result does not depend on k so the distribution of k does not affect the final result. For the first integral we have:

$$\int_{\mathbb{R}^2} 1 - \frac{1}{\left[1+\omega_1 l(x+\tau k_i, d)\right]^2}dx = 2\pi\omega^{2/\alpha}\int_0^\infty \frac{1+2t^\alpha}{(1+t^\alpha)^2}tdt \tag{C.26}$$

$$= \omega_1^{2/\alpha}\left(1+\frac{2}{\alpha}\right)\tilde{C}. \tag{C.27}$$

For the last step we integrate by parts and $\tilde{C}$ is defined in (5.10). The second integral is (B.20). □

C.2 Proof of Theorem 6.3

We rewrite (6.20) in terms of $\nu(p_r) \triangleq \lambda_s \Delta(p_r) D^2$ to obtain:

$$P_{\text{out,mix}}(R) \leq \left[1 - \frac{\alpha(\nu(p_r) - \nu(0))}{2\nu(0)}\right]\left[1 - e^{-\nu(p_r)}\right] + \frac{\alpha(\nu(p_r) - \nu(0))}{2\nu(0)} \times \left\{1 + \frac{2\nu(p_r)}{\lambda_{in}\phi_0 D^2 + 2\nu(p_r)} - e^{-\nu(p_r)}\left[1 + \nu(p_r)\left(1 + \frac{2-\alpha}{\alpha D}\mathbb{E}_r\left[||r-d||\right]\right)\right]\right\}. \quad \text{(C.28)}$$

Since $\nu(p_r)$ is linear in p_r we can analyze the concavity of the OP with respect to $\nu(p_r)$ instead of p_r. We do this by studying when the second derivative of the OP upper bound (C.28) with respect to ν is negative. After differentiating twice with respect to ν and rearranging the terms we obtain:

$$\frac{d^2 P_{\text{out,mix}}}{d\nu^2} \leq \left\{2\alpha\lambda_{in}\phi_0 D^2 e^{\nu(p_r)} \frac{2\nu(0) + \lambda_{in}\phi_0 D^2}{(2\nu(p_r) + \lambda_{in}\phi_0 D^2)^3} + \frac{\alpha}{2}\left(1 - \frac{2}{\alpha}\right) p_c(\nu) \frac{\mathbb{E}_r[||r-d||]}{D} - \left(\frac{\alpha}{2} p_c(\nu) + \nu(0)\right)\right\} \frac{e^{-\nu(p_r)}}{\nu(0)}, \quad \text{(C.29)}$$

with:

$$p_c(\nu) = \nu^2 - (\nu(0) + 4)\nu + 2(1 + \nu(0)). \quad \text{(C.30)}$$

We study the derivative in the interval $0 \leq p_r \leq 1$ which maps to the interval $\nu(0) = \lambda_s \delta D^2 \leq \nu \leq \nu(1) = \lambda_s \delta D^2 (1 + 2/\alpha)$. Using standard arguments it is straightforward to show that for $\alpha > 2$ we have $p_c(\nu) > 0$ in this interval whenever:

$$\nu(0) = \lambda_s \delta D^2 \leq \frac{3 - \sqrt{5}}{2}. \quad \text{(C.31)}$$

In addition, using that $\nu(p_r) \geq \nu(0) = \lambda_s \delta D^2 > 0$ we bound:

$$\lambda_{in}\phi_0 D^2 \frac{2\nu(0) + \lambda_{in}\phi_0 D^2}{(2\nu(p_r) + \lambda_{in}\phi_0 D^2)^3} \leq \frac{1}{\lambda_{in}\phi_0 D^2}, \quad \text{(C.32)}$$

in (C.29), to obtain:

$$\frac{d^2 P_{\text{out,mix}}}{d\nu^2} \leq \frac{e^{-\nu(p_r)}}{\nu(0)} \left\{\frac{4\pi\alpha e^{\nu(p_r)}\sigma_{in}^2}{\phi_0 D^2} + \left(\frac{\alpha}{2} - 1\right) p_c(\nu) \frac{\mathbb{E}_r[||r-d||]}{D} - \left(\frac{\alpha}{2} p_c(\nu) + \nu(0)\right)\right\}, \quad \text{(C.33)}$$

where we have used also that $\sigma_{in} = (2\pi\lambda_{in})^{-1}$. It is clear that the first term in (C.29) is positive and that under (C.31) the second term also is, and the third one is negative. Notice also that for each ϕ_0, D, $\alpha > 2$ as $\sigma_{in} \to 0$, the first term goes to zero and $\mathbb{E}_r[||r-d||]/D \to 1$. Now using standard continuity arguments it is straightforward to show that for each ϕ_0, D, $\alpha > 2$ and $\nu(0) = \lambda_s \delta D^2$ satisfying (C.31) if σ_{in} is small enough then the third negative term will be greater that the other two, and hence the second derivative will become negative.

C.3 Proof of Corollary 6.5

In order to find the value of σ_c we should find the smallest root of the second derivative of $P_{\text{out,mix}}$. Since this cannot be done in closed form, we can find an approximate condition for concavity by finding the smallest root in σ_{in} of the upper bound (C.33) of the second derivative of $P_{\text{out,mix}}$. For the high reliability regime we can further approximate $\nu(0) = \lambda_s \delta D^2 \approx 0$ which leads to (6.25). To obtain condition (6.26) we upper bound $\mathbb{E}[||r-d||]$ using (6.23) in (6.25), to obtain:

$$2\alpha s^2 + (\alpha - 2)\gamma(s, \phi_0)s - 2 = 0. \tag{C.34}$$

This equation cannot be solved in closed form either due to the presence of the function γ, but it can be shown that an absolute upper bound the smallest root is obtained from setting $\alpha = 2$ in the equation. In that case the equation is independent of γ and can be solved in closed form to obtain the condition $s \leq 1/\sqrt{2}$. In can be shown that setting $s = 1/\sqrt{2}$ in $\gamma(s, \phi_0)$ and solving (C.34) yields a lower bound on the smallest root for each value of α. Thus, (C.34) becomes a second degree polynomial in s which can be solved in closed form to obtain (6.26).

C.4 Proof of Theorem 6.6

In Theorem 6.3 we showed that for each network setup such that (C.31) holds there is an interval $\sigma_{in} \leq \sigma_c$ in which the OP upper bound is concave in p_r. Now we show that under this condition, there is an interval in which $p_r = 1$ is optimal by finding conditions such that $P_{\text{out,mix}}(p_r = 1) - P_{\text{out,mix}}(p_r = 0) \leq 0$. Setting $p_r = 0$ and $p_r = 1$ we can write:

$$\begin{aligned} P_{\text{out,mix}}(p_r = 1) - P_{\text{out,mix}}(p_r = 0) \leq e^{-\lambda_s \delta D^2} + \frac{2\lambda_s \Delta}{2\lambda_s \Delta + \phi_0 \lambda_{in}} \\ - \left[1 + \lambda_s \Delta D^2 \left(1 + \frac{2-\alpha}{\alpha D} \mathbb{E}_r \left[||r-d||\right]\right)\right] e^{-\lambda_s \Delta D^2}, \end{aligned} \tag{C.35}$$

where we take $\Delta \equiv \Delta(1) = \delta\left(1 + \frac{2}{\alpha}\right)$. Now we upper bound:

$$\frac{2\lambda_s \Delta}{2\lambda_s \Delta + \phi_0 \lambda_{in}} \leq \frac{2\lambda_s \Delta}{\phi_0 \lambda_{in}} = \frac{4\pi \lambda_s \Delta \sigma_{in}^2}{\phi_0}, \tag{C.36}$$

and:

$$e^{-\lambda_s \delta D^2} - e^{-\lambda_s \Delta D^2}\left(1 + \lambda_s \Delta D^2\right) \leq \lambda_s \delta D^2 e^{-\lambda_s \Delta D^2}\left(\frac{4}{\alpha^2}\lambda_s \delta D^2 - 1\right), \tag{C.37}$$

which is valid when (C.31) is met since then $e^{\nu(0)} \leq 1 + \nu(0) + \nu^2(0)$. With this we obtain:

$$P_{\text{out,mix}}(p_r = 1) - P_{\text{out,mix}}(p_r = 0) \leq \lambda_s \delta D^2 e^{-\lambda_s \Delta D^2} \times \left\{ \left[1 + \frac{2}{\alpha} \right] \left[\frac{4\pi\sigma_{in}^2}{\phi_0 D^2} e^{\lambda_s \Delta D^2} + \left(1 - \frac{2}{\alpha} \right) \frac{\mathbb{E}_r[||r - d||]}{D} \right] + \frac{4\lambda_s \delta D^2}{\alpha^2} - 1 \right\}. \quad \text{(C.38)}$$

Continuity arguments similar to those of Theorem 6.3 prove that if σ_{in} is small enough then the right side of (C.38) will be negative and $p_r = 1$ will be optimal. To find and estimate for the maximum value of σ_{in} we can find the roots of the right side of this expression, focusing on the terms between brackets. In the high reliability regime the term $\lambda_s \delta D^2$ will be small (as shown at the end of Section II) so an approximate condition for concavity can be obtained by letting $\lambda_s \delta D^2 \to 0$, which leads to (6.29). The proof of the simpler condition (6.30) is obtained following the same arguments as in the proof of corollary 6.5, except that in this case we can establish the condition $s \leq 1/2$.

C.5 Proof of Theorem 6.7

When $p_r = 1$ is optimal we can compute the gains starting from (C.38), valid under (C.31), and noting that $P_{\text{out,mix}}(p_r = 0) = P_{\text{out,DT}}$. Rearranging the terms we obtain:

$$\frac{P_{\text{out,mix}}}{P_{\text{out,DT}}} \leq 1 + \frac{\lambda_s \delta D^2}{P_{\text{out,DT}}} \left\{ \left[\left(1 - \frac{2}{\alpha} \right) \frac{\mathbb{E}_r[||r - d||]}{D} e^{-\lambda_s \Delta D^2} + \frac{4\pi\sigma_{in}^2}{\phi_0 D^2} \right] \left(1 + \frac{2}{\alpha} \right) + e^{-\lambda_s \Delta D^2} \left(\frac{4\lambda_s \delta D^2}{\alpha^2} - 1 \right) \right\}. \quad \text{(C.39)}$$

By noting that when $p_r = 1$ is optimal the term between brackets in the previous expression will be negative, we can upper bound this by removing the term $(\lambda_s \delta D^2)/P_{\text{out,DT}}$ outside the brackets. To simplify the expression for the high reliability regime, we can take the approximation $\lambda_s \delta D^2 \approx 0$. Finally, when $p_r = 0$ the performance will be the same as DT so the gain will be one.

Appendix D

Proofs of Chapter 7

D.1 Proof of Lemma 7.1

For shortness we write $N_s(K)$, and using the Campbell-Mecke formula for an independently marked point process (4.22) we have:

$$\mathbb{E}\left[N_s(K)\right] = \lambda_p \int_{\mathbb{R}^2} \mathbb{E}^x \left[\sum_{i=1}^{N_{x,r}} \mathbb{1}_{\{x+D_{x,i}\in K\}} \mathbb{1}_{\mathcal{S}_{x,i}} \right] dx, \tag{D.1}$$

where $\mathbb{E}^x$ averages with respect to the Palm probability of a point process having a cluster at x. With all this we can write:

$$\mathbb{E}\left[N_s(K)\right] = \lambda_p \int_{\mathbb{R}^2} \mathbb{E}^0 \left[\sum_{i=1}^{N_{0,r}} \mathbb{1}_{\{D_{0,i}\in K_{-x}\}} \mathbb{1}_{\mathcal{S}_{0,i}} \right] dx, \tag{D.2}$$

where K_{-x} is the set obtained by shift every point in K by $(-x)$. Moving the integral inside we conclude by noting that:

$$\int_{\mathbb{R}^2} \mathbb{1}_{\{D_{0,i}\in K_{-x}\}} dx = |K|. \tag{D.3}$$

D.2 Proof of Lemma 7.2

Suppose a protocol $(\Phi_p, \mathcal{F})$ is fixed. We can write $\mathcal{S}_{0,i} \equiv \mathcal{S}_{0,i}(R)$ to indicate the dependence of the service event with the required rate R. For a given realization of $\tilde{\Phi}$, decreasing R cannot decrease the number of served requests, that is:

$$\sum_{i=0}^{N_{0,r}} \mathbb{1}_{\mathcal{S}_{0,i}(R_1)} \leq \sum_{i=0}^{N_{0,r}} \mathbb{1}_{S_{0,i}(R_2)}, \tag{D.4}$$

when $R_1 > R_2$. Taking expectation $\mathbb{E}^0[\cdot]$ on both sides, we can upper bound the right by taking the limit as $R_2 \to 0$. But when $R = 0$ we have that: $\mathcal{S}_{0,i} = \mathcal{M}_{0,i}$, that is, when the

required rate is 0, a request is served whenever the file is available in the cluster, and hence, applying the Monotone Convergence Theorem [92]:

$$\mathbb{E}^0\left[\sum_{i=0}^{N_{0,r}} \mathbb{1}_{S_{0,i}(R)}\right] \leq \mathbb{E}^0\left[\sum_{i=0}^{N_{0,r}} \mathbb{1}_{\mathcal{M}_{0,i}}\right] = \mathbb{E}^0[N_{0,r}]p_{\mathcal{M}},$$

with equality as $R \to 0$. To derive the match probability we take the following steps:

$$\begin{aligned}
p_{\mathcal{M}} &= \mathbb{P}^0\left(\cup_{j=1}^{N_{0,u}}\{V_{0,i} \in \mathbf{A}_{0,j}\}\right) && \text{(D.5)}\\
&= \mathbb{E}^0\left[1 - \mathbb{P}^0\left(\cap_{j=1}^{N_{0,u}}\{V_{0,i} \notin \mathbf{A}_{0,j}\}|V_{0,i}, N_{0,u}\right)\right] && \text{(D.6)}\\
&= 1 - \mathbb{E}^0\left[\mathbb{P}^0\left(\cap_{j=1}^{N_{0,u}}\{V_{0,i} \notin \mathbf{A}_{0,j}\}|V_{0,i}, N_{0,u}\right)\right] && \text{(D.7)}\\
&= 1 - \mathbb{E}^0\left[\mathbb{P}^0\left(V_{0,i} \notin \mathbf{A}_{0,j}|V_{0,i}\right)^{N_{0,u}}\right] && \text{(D.8)}\\
&= 1 - \mathbb{E}^0\left[\mathbb{P}^0\left(V_{0,i} \neq A_{0,1,1}|V_{0,i}\right)^{MN_{0,u}}\right] && \text{(D.9)}\\
&= 1 - \mathbb{E}^0\left[\left(1 - p_A(V_{0,i})\right)^{MN_{0,u}}\right] && \text{(D.10)}\\
&= 1 - \mathbb{E}^0\left[e^{-\lambda_u \pi R_c^2[1-(1-p_A(V))^M]}\right]. && \text{(D.11)}
\end{aligned}$$

In the last step we used that $N_{0,u}$ is a Poisson random variable of mean $\lambda_u \pi R_c^2$, independent of V.

D.3 Proof of Lemma 7.3

It is straightforward to show that, given two constants $A, B > 0$, the function

$$\phi(u) = \mathsf{C}\left(\frac{A}{B+u}\right) \tag{D.12}$$

is convex for $u \geq 0$. Now, we rewrite the rate (7.33) in a way suitable to use the convexity of $\phi(u)$. We can now write the achievable rate (7.30) in terms of the interferences (7.31) and (7.32) as:

$$\frac{1}{\Delta}\sum_{i=1}^{\Delta/n_1} \mathsf{C}\left(\mathrm{SIR}_i(x,y)\right) = \frac{1}{n_1}\frac{1}{2}\sum_{u_1=1}^{2}\cdots\frac{1}{2}\sum_{u_{\hat{k}}=1}^{2}\mathsf{C}\left(\frac{|h_{xy}|^2 l(x,y)}{I_b + I_1(u_1) + \ldots + I_{u_{\hat{k}}}(u_1,\ldots u_{\hat{k}})}\right). \tag{D.13}$$

where for shortness we defined $\hat{k} = \log(\Delta/n_1)$. This expression consists of $\hat{k}$ nested convex combinations with two terms each, giving the Δ/n_1 terms seen in (7.30). Notice that the k-th interference term, $I_k(u_1,\ldots,u_k)$, depends only on the summations over $(u_1,\ldots,u_k)$ and is constant for indexes $(u_{k+1},\ldots,u_{\hat{k}})$, because interference from clusters with less slots will change less often. We can therefore recursively use the convexity of $\phi(u)$ to transfer the summations inside $\mathcal{C}(\cdot)$, innermost to outermost. We take $A = |h_{xy}|^2 l(x,y)$ and

$B(k,u_1,\ldots,u_k) = I_b + \sum_{i=1}^{k} I_i(u_1,\ldots,u_i)$. In the first step, for any set of indexes $u_1,\ldots,u_{\hat{k}-1}$ we can write:

$$\frac{1}{2}\sum_{u_{\hat{k}}=1}^{2} \mathsf{C}\left(\frac{|h_{xy}|^2 l(x,y)}{I_b + I_1(u_1) + \ldots + I_{u_{\hat{k}}}(u_1,\ldots u_{\hat{k}})}\right) = \frac{1}{2}\sum_{u_{\hat{k}}=1}^{2} \mathsf{C}\left(\frac{A}{B(\hat{k}-1,u_1,\ldots,u_{\hat{k}-1}) + I_{u_{\hat{k}}}(u_1,\ldots u_{\hat{k}})}\right)$$
$$\geq \mathsf{C}\left(\frac{|h_{xy}|^2 l(x,y)}{B(\hat{k}-1,u_1,\ldots,u_{\hat{k}-1}) + \frac{1}{2}\sum_{u_{\hat{k}}=1}^{2} I_{u_{\hat{k}}}(u_1,\ldots u_{\hat{k}})}\right).$$

Continuing this procedure recursively until all the summations are transfered inside C, we conclude the proof.

D.4 Proof of Theorem 7.4

We focus on the typical cluster at the origin, removing all subindexes 0 for shortness (that is, $N_{0,r} \equiv N_r$, $\mathcal{T}_{0,i} \equiv \mathcal{T}_i$, etc.). To determine the local metric, we need to find:

$$\mathbb{E}^0\left[\sum_{i=1}^{N_r} \mathbb{1}_{\mathcal{M}_i}\mathbb{1}_{\mathcal{T}_i}\right] = \mathbb{E}^0\left[\sum_{i=1}^{N_r} \mathbb{E}^0\left[\mathbb{1}_{\mathcal{M}_i}\mathbb{1}_{\mathcal{T}_i}|N_r,N_m\right]\right], \tag{D.14}$$

where we have used (7.7). We focus on the innermost expectation, which should not depend on i:

$$\mathbb{E}^0\left[\mathbb{1}_{\mathcal{M}_i}\mathbb{1}_{\mathcal{T}_i}|N_r,N_m\right] = \mathbb{P}^0\left(\mathcal{T}_i|N_r,N_m,\mathcal{M}_i\right)\mathbb{P}^0\left(\mathcal{M}_i|N_r,N_m\right). \tag{D.15}$$

To find $\mathbb{P}^0\left(\mathcal{M}_i|N_r,N_m\right)$ we can write:

$$\mathbb{P}^0\left(\mathcal{M}_i|N_r,N_m\right) = \mathbb{E}^0\left[\mathbb{P}^0\left(\mathcal{M}_i|N_r,N_m,N_u,\tilde{\mathbf{A}}\right)|N_r,N_m\right]. \tag{D.16}$$

Then we have:

$$\mathbb{P}^0\left(\mathcal{M}_i|N_r,N_m,N_u,\tilde{\mathbf{A}}\right) = \mathbb{P}^0\left(V_i \in \tilde{\mathbf{A}}|N_r,N_m,N_u,\tilde{\mathbf{A}}\right) = \frac{\binom{N_r-1}{N_m-1}}{\binom{N_r}{N_m}} = \frac{N_m}{N_r}. \tag{D.17}$$

To prove (D.17) we consider that once $\tilde{\mathbf{A}}$ is fixed, the event $V_i \in \tilde{\mathbf{A}}$ can be interpreted as a success in a Bernoulli trial. Then, we are asking for the probability of a success on the i-th trial out of N_r Bernoulli trials given that there were a total of N_m successes. The binomial coefficient in the denominator is the number of ways to assign the N_m successes in the N_r trials, and the numerator is the number of ways to assign $N_m - 1$ successes among $N_r - 1$ trials, given that the i-th trial is a success. Using (D.17) in (D.16), we have:

$$\mathbb{P}^0\left(\mathcal{M}_i|N_m,N_r\right) = \mathbb{E}^0\left[\frac{N_m}{N_r}\middle|N_r,N_m\right] = \frac{N_m}{N_r}. \tag{D.18}$$

We now only need to find $\mathbb{P}^0\left(\mathcal{T}_i|N_r, N_m, \mathcal{M}_i\right)$ in (D.15). The event $\mathcal{T}_i$ in (D.15) can be written as $\mathcal{T}_i = \mathcal{O}_i(W(N_m)) \cap \mathcal{P}_i$, where $\mathcal{P}_i$ indicates the user was scheduled for a transmission and $\mathcal{O}_i(W(N_m)) = \{\text{Tx successful for user } i\}$. In what follows we omit the dependence of $W(N_m)$, $W_L(N_m)$ and $W_H(N_m)$ with N_m. Thus, for $\mathbb{P}^0\left(\mathcal{T}_i|N_r, N_m, \mathcal{M}_i\right)$ in (D.15), we can write:

$$\mathbb{P}^0\left(\mathcal{T}_i|N_r, N_m, \mathcal{M}_i\right) = \mathbb{P}^0\left(\mathcal{O}_i(W), \mathcal{P}_i, W = W_H|N_r, N_m, \mathcal{M}_i\right) + \mathbb{P}^0\left(\mathcal{O}_i(T), \mathcal{P}_i, W = W_L|N_r, N_m, \mathcal{M}_i\right). \quad \text{(D.19)}$$

When $W(N_m) = W_H(N_m)$ the user is always scheduled, which means that $\{W(N_m) = W_H(N_m)\} \subset \mathcal{P}_i$. Then we have:

$$\mathbb{P}^0\left(\mathcal{O}_i(W), \mathcal{P}_i, W = W_H|N_r, N_m, \mathcal{M}_i\right) = \mathbb{P}^0\left(\mathcal{O}_i(W), W = W_H|N_r, N_m, \mathcal{M}_i\right) \quad \text{(D.20)}$$

$$= \mathbb{P}^0\left(\mathcal{O}_i(W_H)|N_r, N_m, \mathcal{M}_i\right) \mathbb{1}_{\{W=W_H\}}. \quad \text{(D.21)}$$

When $\{W(N_m) = W_L(N_m)\}$ the user may not be scheduled so applying Bayes' rule we have:

$$\mathbb{P}^0\left(\mathcal{O}_i(W), \mathcal{P}_i, W = W_L|N_r, N_m, \mathcal{M}_i\right) = \mathbb{P}^0\left(\mathcal{O}_i(W_L)|\mathcal{P}_i, N_r, N_m, \mathcal{M}_i\right) \times \mathbb{P}^0\left(\mathcal{P}_i|N_r, N_m, \mathcal{M}_i\right) \mathbb{1}_{\{W=W_L\}}. \quad \text{(D.22)}$$

By a similar argument to the one used in (D.17) we have:

$$\mathbb{P}^0\left(\mathcal{P}_i|N_r, N_m, \mathcal{M}_i\right) = \frac{W_L(N_m)}{N_m}. \quad \text{(D.23)}$$

Now we need to find the probability of $\mathcal{O}_i(W)$ appearing in (D.21) and (D.22):

$$\mathbb{P}^0\left(\mathcal{O}_i(W_L)|\mathcal{P}_i, N_r, N_m, \mathcal{M}_i\right) = \mathbb{E}^0\left[\sum_{V_i \in \tilde{\mathbf{A}}} \mathbb{P}^0(\mathcal{O}_i(W_L), V_i|\mathcal{P}_i, N_r, N_m, N_u, \mathcal{M}_i, \tilde{\mathbf{A}})|\mathcal{P}_i, N_r, N_m, \mathcal{M}_i\right]. \quad \text{(D.24)}$$

Notice that, we only need to add over the $V_i \in \tilde{\mathbf{A}}$ because we are conditioning on $\mathcal{M}_i$. This guarantees that in the following step we will not condition with respect to an event of zero probability. Now, since there is a match and only one transmitter will serve the request we have:

$$\begin{aligned}\mathbb{P}^0(\mathcal{O}_i(W_L)|V_i, \mathcal{P}_i, N_r, N_m, N_u, \mathcal{M}_i, \tilde{\mathbf{A}}) &= \sum_{c \in \mathcal{A}(\tilde{\mathbf{A}}, V_i)} \mathbb{P}^0(\mathcal{O}_i(W_L), C_i = c|V_i, \mathcal{P}_i, N_r, N_m, N_u, \mathcal{M}_i, \tilde{\mathbf{A}}) \\ &= \sum_{c \in \mathcal{A}(\tilde{\mathbf{A}}, V_i)} \frac{\mathbb{P}^0(\mathcal{O}_i(W_L)|C_i = c, V_i, \mathcal{P}_i, N_r, N_m, N_u, \mathcal{M}_i, \tilde{\mathbf{A}})}{\#\mathcal{A}(\tilde{\mathbf{A}}, V_i)} \\ &= \mathbb{P}^0(R_a(W_L, S, D_i) > R|N_r, N_m, N_u), \end{aligned} \quad \text{(D.25)}$$

where S follows the distribution of any user in the cluster (uniform), R_a is given by (7.33), and # denotes the number of elements in the set. In the last step we use that the event of failed transmission to a user depends only on the number of slots in the block (that is, on N_m) and on the number of requests and caching users (otherwise the transmission may not be well defined). Also, we use that it does not depend on the value of C_i, or on other variables such as the video being transmitted (this is a consequence of the factorization of the distribution of the cluster random variables (7.7) and the definition of the in-cluster communication strategy, Sec. 7.3.1).

Going back to (D.24), applying Bayes' rule, and replacing (D.25) we get:

$$\begin{aligned}
&\mathbb{P}^0\left(\mathcal{O}_i(W_L)|\mathcal{P}_i, N_r, N_m, \mathcal{M}_i\right) \\
&= \mathbb{E}^0\left[\mathbb{P}^0(R_a(W_L, S, D_i) < R|N_r, N_m, N_u) \sum_{V_i \in \tilde{\mathbf{A}}} \mathbb{P}^0(V_i|\mathcal{P}_i, N_r, N_m, N_u, \mathcal{M}_i, \tilde{\mathbf{A}})|N_r, N_m\right], \\
&= \mathbb{P}^0(R_a(W_L, S, D_i) > R|N_r, N_m). \quad \text{(D.26)}
\end{aligned}$$

The sum above is one, because we are conditioning on a match and on $\tilde{\mathbf{A}}$. Continuing from (D.26) we get:

$$\begin{aligned}
\mathbb{P}^0\left(\mathcal{O}_i(W_L)|\mathcal{P}_i, N_r, N_m, \mathcal{M}_i\right) &= \mathbb{P}^0(|h_{S,D_i}|^2 l(S, D_i) > I(D_i, N_m) W_L(N_m) R|N_r, N_m) \quad \text{(D.27)} \\
&= \mathbb{E}^0\left[\bar{F}_{|h|^2}\left(\frac{I(D_i, N_m) W_L(N_m) R}{l(S, D_i)}\right)\middle| N_r, N_m\right], \quad \text{(D.28)}
\end{aligned}$$

where $\bar{F}_{|h|^2}$ is the complementary distribution function of the fading random variable and $I(D_i, N_m)$ is the interference at D_i (7.34). Following the same procedure in (D.21) with $\mathbb{P}^0\left(\mathcal{O}_i(W_H)|N_r, N_m, \mathcal{M}_i\right)$ we can find a formula identical to (D.28) but replacing W_L by W_H. Then, using these two formulae and (D.17) in (D.21) and (D.22), we can find (D.19):

$$\begin{aligned}
\mathbb{P}^0\left(\mathcal{T}_i|N_r, N_m, \mathcal{M}_i\right) &= \mathbb{E}^0\left[\sum_{i=1}^{N_r} \mathbb{E}^0\left[\mathbb{1}_{\mathcal{M}_i} \mathbb{1}_{\mathcal{T}_i}|N_r, N_m\right]\right], \\
&= \mathbb{E}^0\left[\bar{F}_{|h|^2}\left(\frac{I(D, N_m) W(N_m) R}{l(S, D)}\right)\middle| N_r, N_m\right] \\
&\quad \times \left(\mathbb{1}_{\{W(N_m) = W_H(N_m)\}} + \frac{W_L(N_m)}{N_m} \mathbb{1}_{\{W(N_m) = W_L(N_m)\}}\right). \quad \text{(D.29)}
\end{aligned}$$

Finally, using (D.17) and (D.29) in (D.14) we get the desired result:

$$\begin{aligned}
\mathbb{E}^0\left[\sum_{i=1}^{N_r} \mathbb{E}^0\left[\mathbb{1}_{\mathcal{M}_i} \mathbb{1}_{\mathcal{T}_i}|N_r, N_m\right]\right] &= \mathbb{E}^0\left[N_r \mathbb{E}^0\left[\bar{F}_{|h|^2}\left(\frac{I(D, N_m) W(N_m) R}{l(S, D)}\right)\middle| N_r, N_m\right]\right. \\
&\quad \left. \times \left(\mathbb{1}_{\{W(N_m) = W_H(N_m)\}} + \frac{W_L(N_m)}{N_m} \mathbb{1}_{\{W(N_m) = W_L(N_m)\}}\right) \frac{N_m}{N_r}\right], \quad \text{(D.30)}
\end{aligned}$$

where we have replaced D_i by another random variable D with the same distribution as D_i. For unit mean Rayleigh fading $\bar{F}_{|h_{S,D}|^2}(u) = \exp(-u)$.

The proof of the average rate follows on the same lines:

$$\bar{R}(\mathcal{F}, R) = R\ \mathbb{E}^0\left[\sum_{i=1}^{N_r} \frac{\mathbb{1}_{\mathcal{S}_i}}{\sum_{j=1}^{N_r} \mathbb{1}_{\mathcal{M}_j \cap \mathcal{P}_j}}\right] \tag{D.31}$$

$$= R\ \mathbb{E}^0\left[\sum_{i=1}^{N_r} \frac{\mathbb{1}_{\mathcal{M}_i \cap \mathcal{P}_i} \mathbb{E}^0[\mathbb{1}_{\mathcal{O}_i} | \mathcal{M}_i, \mathcal{P}_i, N_r, N_m]}{\sum_{j=1}^{N_r} \mathbb{1}_{\mathcal{M}_j \cap \mathcal{P}_j}}\right]. \tag{D.32}$$

In the last step we can use (D.28) with $W(N_m)$ instead of $W_L(N_m)$, and see that the conditional expectation does not depend on i, which concludes the proof.

D.5 Proof of Lemma 7.5

We review the notion of second-order stochastic dominance [93]. If X, Y are real random variables, then X second-order stochastically dominates Y, written $X \geq_2 Y$ if, for all $u > 0$:

$$\int_0^u F_X(x)dx \leq \int_0^u F_Y(y)dy.$$

This definition is equivalent to asking that for any monotone increasing concave function $\phi(u)$:

$$\mathbb{E}[\phi(X)] \geq \mathbb{E}[\phi(Y)]. \tag{D.33}$$

The following theorems are useful [94]:

- If $\{X_1, \dots X_n\}$ and $\{Y_1, \dots, Y_n\}$ are sets of independent real random variables and $\{\beta_i\}_{i=1}^n$ are non-negative real numbers:

$$X_i \geq_2 Y_i\ \forall i \iff \sum_{i=1}^n \beta_i X_i \geq_2 \sum_{i=1}^n \beta_i Y_i. \tag{D.34}$$

- If $\{X_1, \dots X_n\}$ are independent real random variables and $\{\beta_i\}_{i=1}^n$ are non-negative real numbers such that $\sum_{i=1}^n \beta_i = 1$:

$$\frac{1}{n}\sum_{i=1}^n X_i \geq_2 \sum_{i=1}^n \beta_i X_i. \tag{D.35}$$

We now consider the interference (7.34) generated only by the clusters with centers inside $\mathcal{B}(0, \rho)$:

$$I_\rho(d, n_1) = \sum_{x \in (\Phi_p \cap \mathcal{B}(0,\rho)) \setminus \{0\}} \psi(x, d, \mathbf{m}_x, n_1), \tag{D.36}$$

where $\psi(x, d, \mathbf{m}_x, n_1)$ is the function in the sum in (7.34). This sum has a finite number of terms almost surely. Additionally, let $\tilde{\psi}(x, d, \mathbf{m}_x, n_1)$ be given by (7.41), that is, be identical

to $\psi(x, d, \mathbf{m}_x, n_1)$ but with all the $\{\mathcal{B}_{x,i,j}\}$ equal to 1. Now, we condition on the cluster centers Φ_p, on the number of slots $\{W_x\}$ and on the variables $\{B_{x,i,j}\}$ of all the clusters. Then the only randomness in the sum (D.36) comes from the fading coefficients and the positions of the users around the cluster centers, which are all independent random variables. With this, for a cluster with $W_x = 2^i > n_1$ and using (7.35), (D.35) we have:

$$\psi(x, d, \mathbf{m}_x, n_1) = \frac{n_1}{2^i} \sum_{j=1}^{2^i/n_1} B_{x,j,i} |h_{x,j,d}|^2 l(x + r_j, d) \tag{D.37}$$

$$\leq_2 \frac{n_1}{2^i} \sum_{j=1}^{2^i/n_1} |h_{x,j,d}|^2 l(x + r_j, d) \tag{D.38}$$

$$= \tilde{\psi}(x, d, \mathbf{m}_x, n_1). \tag{D.39}$$

Keeping the conditioning, and summing over all the clusters which have more that n_1 slots and using (D.34):

$$\sum_{x \in (\Phi_p \cap \mathcal{B}(0,\rho)) \setminus \{0\}} \tilde{\psi}(x, d, \mathbf{m}_x, n_1) \mathbb{1}_{\{W_x > n_1\}} \geq_2 \sum_{x \in (\Phi_p \cap \mathcal{B}(0,\rho)) \setminus \{0\}} \psi(x, d, \mathbf{m}_x, n_1) \mathbb{1}_{\{W_x > n_1\}}. \tag{D.40}$$

With the conditioning, the interference from clusters such that $W_x \leq n_1$ is independent of the ones such that $W_x > n_1$, so using (D.34) once more get:

$$\sum_{x \in (\Phi_p \cap \mathcal{B}(\rho)) \setminus \{0\}} \tilde{\psi}(x, d, \mathbf{m}_x, n_1) \geq_2 I_\rho(d, n_1). \tag{D.41}$$

This means we can use (D.33) with $\phi(u) = -e^{-su}$, $s > 0$, which is concave and increasing in u. Averaging over the conditioned random variables we get the desired result but for $I_\rho(d, n_1)$ instead of the full interference. The proof concludes by letting $\rho \to \infty$ and using monotonicity arguments.

Bibliography

[1] CISCO Systems, "The Zettabyte Era. Trends and analysis," *White paper*, June 2014.

[2] D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic Geometry and its Applications*, 2nd ed. Wiley, 2009.

[3] A. Goldsmith, *Wireless Communications*. New York, NY, USA: Cambridge University Press, 2005.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory, 2nd Edition*. New York: Wiley-Interscience, 2006.

[5] C. Shannon, "A mathematical theory of communication," *Bell System Technical Journal, The*, vol. 27, no. 3, pp. 379–423, July 1948.

[6] B. Picinbono, "On circularity," *IEEE Trans. Signal Process.*, vol. 42, no. 12, pp. 3473–3482, Dec. 1994.

[7] M. Haenggi, *Stochastic Geometry for Wireless Networks*. Cambridge University Press, 2012.

[8] F. Baccelli and B. Błaszczyszyn, *Stochastic Geometry and Wireless Networks*. NoW Publishers, 2009.

[9] M. Haenggi and R. K. Ganti, *Interference in Large Wireless Networks*. NoW Publishers, 2010.

[10] F. Baccelli, B. Błaszczyszyn, and P. Muhlethaler, "An aloha protocol for multihop mobile wireless networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 421–436, Feb. 2006.

[11] A. Nosratinia, T. Hunter, and A. Hedayat, "Cooperative communication in wireless networks," *Communications Magazine, IEEE*, vol. 42, no. 10, pp. 74–80, Oct 2004.

[12] M. Dohler and Y. Li, *Cooperative communications : hardware, channel & PHY*. Chichester, West Sussex, U.K., Hoboken, NJ: Wiley, 2010.

[13] E. Van Der Meulen, "Three-terminal communication channels," *Adv. Appl. Prob.*, vol. 3, pp. 120–154, 1971.

[14] T. Cover and A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 572–584, Sep. 1979.

[15] A. Høst Madsen and J. Zhang, "Capacity bounds and power allocation for wireless relay channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2020–2040, Jun. 2005.

[16] J. N. Laneman, D. N. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062– 3080, Dec. 2004.

[17] V. Chandrasekhar, J. Andrews, and A. Gatherer, "Femtocell networks: a survey," *Communications Magazine, IEEE*, vol. 46, no. 9, pp. 59 –67, september 2008.

[18] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *preprint arXiv:1310.0720*, 2013.

[19] X. Lin, J. G. Andrews, A. Ghosh, and R. Ratasuk, "An overview of 3GPP device-to-device proximity services," *IEEE Communications Magazine*, December 2013.

[20] M. Ji, G. Caire, and A. F. Molisch, "The throughput-outage tradeoff of wireless one-hop caching networks," *preprint arXiv:1312.2637*, 2013.

[21] J. G. Proakis, *Digital Communications*, 5th ed. Mc Graw Hill, 2008.

[22] J. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. Soong, and J. Zhang, "What will 5G be?" *Selected Areas in Communications, IEEE Journal on*, vol. 32, no. 6, pp. 1065–1082, June 2014.

[23] A. E. Gamal and Y. Kim, *Network Information Theory*. Cambridge University Press, 2012.

[24] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[25] X. Lin, J. Andrews, A. Ghosh, and R. Ratasuk, "An overview of 3gpp device-to-device proximity services," *Communications Magazine, IEEE*, vol. 52, no. 4, pp. 40–48, April 2014.

[26] A. J. Baddeley, *Lecture Notes in Mathematics: Stochastic Geometry*, 2nd ed. Springer Berlin Heidelberg, 2007.

[27] D. J. Daley and D. Vere-Jones, *An introduction to the theory of point processes. Vol. I*, 2nd ed. New York: Springer, 2008.

[28] ——, *An introduction to the theory of point processes. Vol. II*, 2nd ed. New York: Springer, 2008.

[29] R. Schneider and W. Weil, *Stochastic and Integral Geometry*, 1st ed. Springer, 2008.

[30] M. Haenggi, J. G. Andrews, F. Baccelli, O. Dousse, and M. Franceschetti, "Stochastic geometry and random graphs for the analysis and design of wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 7, pp. 1029–1046, Sep. 2009.

[31] H. ElSawy, E. Hossain, and M. Haenggi, "Stochastic Geometry for Modeling, Analysis, and Design of Multi-tier and Cognitive Cellular Wireless Networks: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, pp. 996–1019, Jul. 2013.

[32] J. F. C. Kingman, *Poisson Processes*. Oxford University Press, 1993.

[33] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.

[34] S. Parkvall, A. Furuskar, and E. Dahlman, "Evolution of LTE toward IMT-advanced," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 84–91, 2011.

[35] Rohde & Schwarz, "LTE- advanced (3GPP Rel.11) technology introduction," *White Paper*, Jul 2013.

[36] R. Ganti and M. Haenggi, "Spatial and temporal correlation of the interference in aloha ad hoc networks," *Communications Letters, IEEE*, vol. 13, no. 9, pp. 631–633, Sept 2009.

[37] M. Haenggi and R. Smarandache, "Diversity polynomials for the analysis of temporal correlations in wireless networks," *Wireless Communications, IEEE Transactions on*, vol. 12, no. 11, pp. 5940–5951, November 2013.

[38] M. Katz and S. Shamai, "Cooperative schemes for a source and an occasional nearby relay in wireless networks," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5138–5160, Nov. 2009.

[39] Y. Zhu, Y. Xin, and P.-Y. Kam, "Optimal transmission strategies for Rayleigh fading relay channels," *IEEE Trans. Wireless Commun.*, vol. 7, no. 2, pp. 618–628, Feb 2008.

[40] L. Zhang, J. Jiang, A. Goldsmith, and S. Cui, "Study of Gaussian relay channels with correlated noises," *IEEE Trans. Commun.*, vol. 59, no. 3, pp. 863–876, March 2011.

[41] D. Skraparlis, V. K. Sakarellos, A. D. Panagopoulos, and J. D. Kanellopoulos, "Outage performance analysis of cooperative diversity with MRC and SC in correlated lognormal channels," *EURASIP Journal on Wireless Comms. and Networking*, vol. 2009, no. 1, pp. 707–839, 2009.

[42] S. Weber, J. G. Andrews, and N. Jindal, "An overview of the transmission capacity of wireless networks," *IEEE Trans. Commun.*, vol. 58, no. 12, pp. 3593–3604, Dec. 2010.

[43] J. Andrews, F. Baccelli, and R. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 3122–3134, 2011.

[44] H. Dhillon, R. Ganti, F. Baccelli, and J. Andrews, "Modeling and analysis of K-tier downlink heterogeneous cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 550 –560, april 2012.

[45] A. Giovanidis and F. Baccelli, "Coverage by pairwise base station cooperation under adaptive geometric policies," *Proc. of the 47th Annual Asilomar*, Nov. 2013.

[46] R. Tanbourgi, S. Singh, J. Andrews, and F. Jondral, "A tractable model for non-coherent joint-transmission base station cooperation," *Wireless Communications, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2014.

[47] R. Heath, M. Kountouris, and T. Bai, "Modeling heterogeneous network interference using poisson point processes," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 4114–4126, 2013.

[48] G. Nigam, P. Minero, and M. Haenggi, "Coordinated Multipoint Joint Transmission in Heterogeneous Networks," *IEEE Transactions on Communications*, 2014, accepted. Available at http://www.nd.edu/~mhaenggi/pubs/tcom14.pdf.

[49] S. P. Weber, X. Yang, J. G. Andrews, and G. de Veciana, "Transmission capacity of wireless ad hoc networks with outage constraints," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4091–4102, Dec. 2005.

[50] A. M. Hunter, J. Andrews, and S. Weber, "Transmission capacity of ad hoc networks with spatial diversity," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5058–5071, Dec. 2008.

[51] N. Jindal, S. Weber, and J. G. Andrews, "Fractional power control for decentralized wireless networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5482–5492, Dec. 2008.

[52] R. Giacomelli, R. Ganti, and M. Haenggi, "Outage probability of general ad hoc networks in the high-reliability regime," *IEEE/ACM Trans. Netw.*, vol. 19, no. 4, pp. 1151 –1163, aug. 2011.

[53] R. K. Ganti, J. G. Andrews, and M. Haenggi, "High-SIR transmission capacity of wireless networks with general fading and node distribution," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3100–3116, May 2011.

[54] A. El Gamal, M. Mohseni, and S. Zahedi, "Bounds on capacity and minimum energy-per-bit for AWGN relay channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1545–1561, Apr. 2006.

[55] A. Carleial, "Multiple-access channels with different generalized feedback signals," *IEEE Trans. Inf. Theory*, vol. 28, no. 6, pp. 841–850, Nov. 1982.

[56] F. Willems and E. van der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders," *IEEE Trans. Inf. Theory*, vol. 31, no. 3, pp. 313–327, May 1985.

[57] C. M. Zeng, F. Kuhlmann, and A. Buzo, "Achievability proof of some multiuser channel coding theorems using backward decoding," *IEEE Trans. Inf. Theory*, vol. 35, no. 6, pp. 1160–1165, Nov. 1989.

[58] A. Behboodi and P. Piantanida, "Selective coding strategy for composite relay channels," in *Communications Control and Signal Processing (ISCCSP), 2012 5th International Symposium on*, May 2012, pp. 1–6.

[59] A. Høst-Madsen, "Cooperation in the low power regime," in *Communication, Control, and Computing (Allerton), 2004 42th Annual Allerton Conference on*, 2004, pp. 1908–1917.

[60] ——, "Capacity bounds for cooperative diversity," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1522–1544, April 2006.

[61] S. H. Lim, Y.-H. Kim, A. El Gamal, and S.-Y. Chung, "Noisy network coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3132–3152, May 2011.

[62] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.

[63] ——, "Towards an information theory of large networks: an achievable rate region," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1877– 1894, Aug. 2003.

[64] L. Xie and P. R. Kumar, "A network information theory for wireless communication: scaling laws and optimal operation," *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 748–767, May 2004.

[65] F. Xue, L. L. Xie, and P. R. Kumar, "The transport capacity of wireless networks over fading channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 834–847, Mar. 2005.

[66] R. K. Ganti and M. Haenggi, "Interference and outage in clustered wireless ad hoc networks," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 4067–4086, Sep. 2009.

[67] H. Q. Nguyen, F. Baccelli, and D. Kofman, "A stochastic geometry analysis of dense IEEE 802.11 networks," in *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE*, 2007, pp. 1199–1207.

[68] S. Weber, J. Andrews, and N. Jindal, "The effect of fading, channel inversion, and threshold scheduling on ad hoc networks," *Information Theory, IEEE Transactions on*, vol. 53, no. 11, pp. 4127 –4149, nov. 2007.

[69] J. Andrews, S. Weber, M. Kountouris, and M. Haenggi, "Random access transport capacity," *IEEE Trans. Wireless Commun.*, vol. 9, no. 6, pp. 2101 –2111, june 2010.

[70] S. Vanka and M. Haenggi, "Analysis of the benefits of superposition coding in random wireless networks," in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, June 2010, pp. 1708–1712.

[71] ——, "Coordinated packet transmission in random wireless networks," in *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE*, Dec 2010, pp. 1–5.

[72] M. Haenggi, "On routing in random Rayleigh fading networks," *Wireless Communications, IEEE Transactions on*, vol. 4, no. 4, pp. 1553 – 1562, july 2005.

[73] ——, "On distances in uniformly random networks," *Information Theory, IEEE Transactions on*, vol. 51, no. 10, pp. 3584 –3586, oct. 2005.

[74] H. Nuttall, "Some integrals involving the q-function," *Naval Underwater Systems Center, New London, CT, Tech. Rep. 4297*, 1972.

[75] R. T. Rockafellar, *Convex Analysis*. Princeton University Press, 1970.

[76] F. Baccelli, A. El Gamal, and D. N. Tse, "Interference networks with Point-to-Point codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2582–2596, May 2011.

[77] N. Golrezaei, A. Molisch, A. Dimakis, and G. Caire, "Femtocaching and device-to-device collaboration: A new architecture for wireless video distribution," *Communications Magazine, IEEE*, vol. 51, no. 4, pp. 142–149, April 2013.

[78] K. Shanmugam, N. Golrezaei, A. Dimakis, A. Molisch, and G. Caire, "Femtocaching: Wireless content delivery through distributed caching helpers," *Information Theory, IEEE Transactions on*, vol. 59, no. 12, pp. 8402–8413, Dec 2013.

[79] N. Golrezaei, A. Dimakis, and A. Molisch, "Wireless device-to-device communications with distributed caching," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, 2012, pp. 2781–2785.

[80] M. Ji, G. Caire, and A. Molisch, "Optimal throughput-outage trade-off in wireless one-hop caching networks," in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, July 2013, pp. 1461–1465.

[81] Q. Ye, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "Resource optimization in device-to-device cellular systems using time-frequency hopping," *IEEE Transactions on Wireless Communications*, 2013, submitted.

[82] ——, "Device-to-device modeling and analysis with a modified Matérn hardcore BS location model," in *Proc. IEEE Globecom*, December 2013.

[83] X. Lin, J. G. Andrews, and A. Ghosh, "Spectrum sharing for device-to-device communication in cellular networks," *IEEE Transactions on Wireless Communications*, 2013, submitted.

[84] B. Matérn, *Spatial Variation*. Springer Lecture Notes in Statistics, 2nd ed., 1986.

[85] M. Haenggi, "Mean interference in hard-core wireless networks," *Communications Letters, IEEE*, vol. 15, no. 8, pp. 792–794, 2011.

[86] A. Altieri, L. Rey Vega, C. G. Galarza, and P. Piantanida, "Analysis of a cooperative strategy for a large decentralized wireless network," *preprint arXiv:1203.3287*, 2012.

[87] A. Altieri, L. Vega, C. Galarza, and P. Piantanida, "Cooperative unicasting in large wireless networks," in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, 2013, pp. 429–433.

[88] A. Khisti, U. Erez, and G. Wornell, "Fundamental limits and scaling behavior of cooperative multicasting in wireless networks," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2762 – 2770, Jun. 2006.

[89] X. Zhang and M. Haenggi, "The Performance of Successive Interference Cancellation in Random Wireless Networks," *IEEE Transactions on Information Theory*, 2013, submitted.

[90] P. Bello, "Characterization of randomly time-variant linear channels," *Communications Systems, IEEE Transactions on*, vol. 11, no. 4, pp. 360–393, December 1963.

[91] J. Marcum, "A statistical theory of target detection by pulsed radar," *Information Theory, IRE Transactions on*, vol. 6, no. 2, pp. 59 –267, april 1960.

[92] R. B. Ash and C. Doléans-Dade, *Probability and Measure Theory*. Harcourt Academic Press, 2000.

[93] E. Wolfstetter, *Topics in microeconomics*. Cambridge Univ. Press, 1999.

[94] C.-K. Li and W.-K. Wong, "Extension of stochastic dominance theory to random variables." *RAIRO - Operations Research*, vol. 33, no. 4, pp. 509–524, 1999.